HAZARDMATCH: AN APPLICATION OF ARTIFICIAL INTELLIGENCE TO LANDSLIDE SUSCEPTIBILITY MAPPING, HOWE SOUND AREA, BRITISH COLUMBIA

Lionel E. Jackson Jr.
Geological Survey of Canada, Pacific Division, 625 Robson Street, Vancouver, BC Canada, V6B5J3
lijackso@nrcan.gc.ca
Clinton P. Smyth
Georeference Online Ltd., Suite 301, 850 West Hastings Street, Vancouver, B.C., Canada, V6C 1E1
David Poole
Department of Computer Science, University of British Columbia, 2366 Main Mall, Vancouver, BC, Canada V6T 1Z4

RÉSUMÉ
HazardMatch est un système informatique conçu pour le dressage de cartes de susceptibilité aux glissements de terrain tout en se servant de la sémantique technologique, un champ d'intelligence artificielle. Ce système fournit un logiciel selon lequel un expert en glissements de terrain peut décrire, en se servant du vocabulaire spécialisé, les caractéristiques des endroits démontrant une haute susceptibilité aux glissements de terrain. Le logiciel se sert du même langage pour générer des descriptions sémantiques de sous-régions d'une région spécifique pour joindre les caractéristiques qui sont constantes. Les cartes de susceptibilité sont donc produites à partir d'un système d'évaluation de ces caractéristiques. L'avantage de cette approche est que les résultats peuvent être facilement expliqués et justifiés à ceux qui ne sont pas experts. De plus, si les résultats sont faux, l'erreur est facilement retraçable à la base de données. Les erreurs peuvent donc être facilement corrigées pour produire des cartes plus précises. Un essai préliminaire de cette méthode a été effectué dans la chaîne côtière à l'est de Howe Sound, Colombie-Britannique.

ABSTRACT
HazardMatch is a computer system for the production of landslide susceptibility maps using semantics and semantic technology, a field of artificial intelligence. It provides a software framework within which a landslide expert can describe, using language as close to natural (specialist) language as possible, the properties of surface locations which are highly susceptible to landslides. It uses the same language to generate semantic descriptions of all sub-areas of the area of interest within which all landslide-relevant properties are effectively constant. By ranking all sub-area descriptions by their similarity to those around extant slope failures, maps of similarity scores and rankings are created. In this way, HazardMatch uses the same sort of knowledge that practitioners use to communicate between themselves in recognizing unstable slope environments. The results are readily explainable and justifiable to non-experts. Furthermore, if the results are demonstrably incorrect, the cause of error is usually traceable to an aspect of the expert's landslide model that is at odds with the facts, or to the input data. These aspects can then be changed to produce more accurate maps. A preliminary test of this method was carried out in the Coast Mountains. Examination of soil slides on recently logged hill slopes yielded promising results.

1. INTRODUCTION

HazardMatch is an internet-based computer system currently under development by Georeference Online Ltd. in collaboration with researchers at the Pacific Division, Geological Survey of Canada and the Department of Computer Science, University of British Columbia. This technology utilizes computer reasoning capabilities to map natural hazard-prone areas by directly using the language of geological and geotechnical practitioners. The last few years have seen a dramatic increase in freely available data that is amenable and relevant to the production of landslide susceptibility maps. Anticipation of this trend was one of the reasons why our group began this research: believing that the availability of large volumes of free data would generate a demand for "intelligent" software that could assist with automating its interpretation. The most relevant Canadian development in this regard was the announcement on 4 April 2007 by Natural Resources Canada of free online access to all its digital mapping data, some of which is relevant to our system.

The reasoning underlying this methodology is similar to that of the experience-base judgment of the geotechnical practitioner. It is intended to be a decision-support tool intended for use by government and community planners, insurance companies and scientific researchers. It takes advantage of the explosion of free or nominally priced sources of geographically based geological and physiographical data available over the Internet. Because the system uses the same sort of knowledge that practitioners use to communicate between themselves, it is inherently auditable i.e. the way it classifies hazards is easier to understand by non-geotechnical experts than other techniques such as multivariate analysis methods e.g.

In : J. Locat, D. Perret, D. Turmel, D. Demers et S. Leroueil, (2008).
Comptes rendus de la 4e Conférence canadienne sur les géorisques: des causes à la gestion.
Proceedings of the 4th Canadian Conference on Geohazards : From Causes to Management. Presse de l'Université Laval, Québec, 594 p.

Guzzetti *et al.* (1999), Guzzetti *et al.* (2006), Suzen, M. L., Doyuran, V. (2006), Wang and Sassa (2007).

This paper will briefly describe the application of the artificial intelligence (AI) technology underlying our system, illustrate its output with the results of a recent practical test, and discuss the next phases of its development.

2. METHODOLOGY

Our system's computational environment is built around knowledge representation technology that has seen considerable advancement over the last decade: most recently in the publication of the Web Ontology Language (OWL) Standard (McGuiness and van Harmelen, 2004). An ontology is essentially a model representing concepts and relationships between them within a domain of knowledge. Classification of rocks, sediments and landslides comprise such a domain. For example, morainal sediments include till and both are composed of granular unconsoidated or semi-consolidated inorganic sediments. Ontologies allow the recognition of these interrelationships and are amenable to processing by a computer so that descriptions assigned to polygons in, for example, bedrock or surfical geology maps can be analyzed for a given purpose. For uniform results to be produced, there must be standardization of terminology in the geologic domain. The Commission for the Management and Application of Geoscience (CGI) published an effective Upper Level Geology Ontology of material relevance to knowledge representation globally. The CGI's geology ontology and its associated standards have been named "GeoSciML" (Geoscience Markup Language), and are documented at:

https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML

The first high-quality commercial OWL ontology editor appeared on the market in 2006 (TopBraid Composer 2006). The first high-quality commercial semantic network storage technology became available in 2007 (AllegroGraph 2007). Our system combines these technologies to create a computational environment within which map polygon descriptors can be compared, along with other geographically-based-variables as slope steepness and orientation, with data models developed for landslide susceptible environments. Areas where the aggregate properties of descriptive and quantitative data are similar to models for landslide occurrence receive high similarity scores whereas areas where the opposite is true receive low similarity scores. These are displayed by colouring polygons according to their similarity scores, thus producing a landslide susceptibility maps where the ratings are directly explicable. These results will be of significant value to land-use planners and others interested in or potentially affected by landslides.

3. WORKFLOW

Figure 1 is a methodology diagram that shows a landslide susceptibility analysis from input to report.

Area 1 represents existing knowledge about landslide susceptibility which is often compiled into "landslide model" descriptions (see next section), some of which are now published on government web sites.

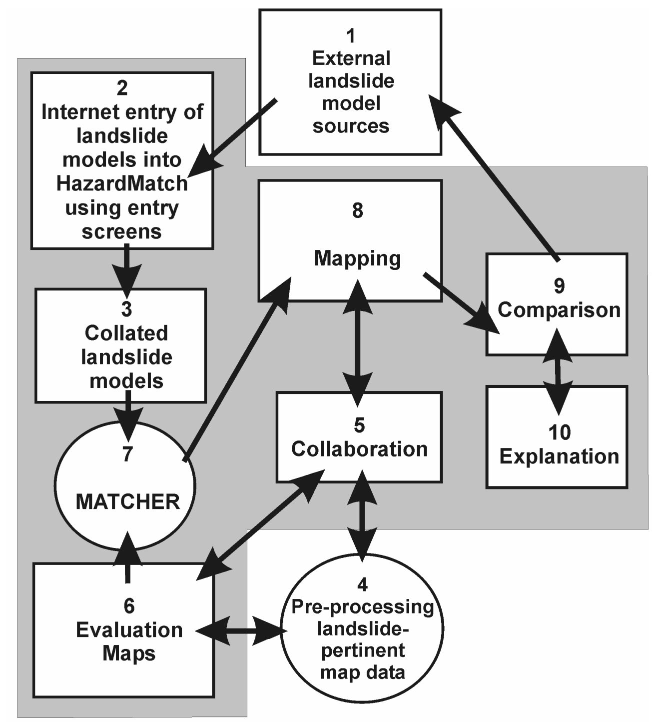

Figure 1. HazardMatch System methodology and workflow. Operations within boxess are dominated by data collection and collation. Circles indicate analytical processing. Areas within the grey background operate within an interactive environment via the internet.

Area 2 depicts 'knowledge capture' screens used for input of landslide models into the system. This is a critical step in the workflow that may involve the translation of terms in the natural language knowledge source to identical or similar terms used by the landslide ontology. As such, it is always undertaken by a geotechnical expert who understands the meanings of the technical terms used to describe landslide susceptibility-related phenomena. Any number of different landslide susceptibility models can be entered into the system at this stage. Each model can produce its own landslide susceptibility maps.

Area 3 represents a collection of landslide susceptibility models ready for comparison with appropriate input maps.

Area 4 represents the environment in which considerable pre-processing of raw landslide-pertinent map data takes place. The pre-processing falls into three categories:

- Terminological – during which terms used in the input maps are made consistent with the (ontology-compliant) terms used in the landslide susceptibility model descriptions of Area 3;
- GIS – where spatial joins are carried out between all relevant polygon, line and point layers to provide individual descriptions of each individually-characterizable polygonal area in the region under study;
- Formatting – during which the representational format of map terminological data is changed from tabular to semantic network format.

Area 5 represents an optional online GIS-collaboration area from which the input data for Area 4 may be sourced, and to which the results from Area 8 (described below) may be recorded.

Area 6 represents a collection of maps from the pre-processing of Area 4 that are ready for comparison with any of the landslide susceptibility models listed in Area 3 or primary data available from Area 5.

Area 7 is the 'Matcher' which lies at the heart of our system. The Matcher's function is to compare a user-selected landslide susceptibility model from the list in Area 3 with all the polygons in a user-selected map from the list in Area 6.

Area 8 represents the environment in which the polygons originating from Area 6 are joined with their similarity scores produced in Area 7 and coloured according to these scores. Usually high scores (indicating high landslide susceptibility in regard to the model being assessed) are given hot colours and low scores cold colours.

Area 9 represents 'comparison reports' produced in pop-up windows by clicking on any polygon in an Area 8 map. This report lists the attributes of the landslide susceptibility model of interest alongside the matching or conflicting attributes of the selected polygon, thus enabling the user to quickly understand how the similarity score was derived.

Area 10 shows 'explanation reports' produced in an additional pop-up window by clicking on any line of interest in the comparison report. This report explains to the user why the model and polygon attributes on any clicked line were paired together for similarity ranking purposes–the justification for which is not always immediately apparent to the non-expert user. Outputs in 9 and 10 can be used to evaluate and modify landslide models.

4. LANDSLIDE MODELS

Our system assumes that the variables associated with various categories of slope failures (e.g. Cruden and Varnes 1996) can be developed into models by appropriately qualified geoscience experts for a given region. These models use natural language concerning tendencies for failures to occur much like an experienced geotechnical practitioner might initially assess a slope under study. For example, based upon his or her experience along a given highway, a practitioner might note that slopes like one under consideration are: ALWAYS unstable, SOMETIMES unstable, RARELY unstable or NEVER unstable. The terms in bold are experience-based tendencies that can be included in our system's reasoning through integration into landslide occurrence models.

To employ our system, one or more occurrence models are constructed for each category of landslide that occurs in an area under study using a standard classification scheme such as Cruden and Varnes (1996). The models were based upon collation of environmental factors associated with each landslide mapped from 1:20 000 air photographs, bedrock lithology (Roddick *et al.* 1979) and surficial geology (Thompson, 1980). The surficial geology was mapped using Terrain Classification System for British Columbia (TCSBC; Howes and Kenk, 1997). TCSBC polygon descriptors are complicated. They contain information about the genesis, texture, landform morphology, and modifying processes. The model is able to access all of this information directly from the designators for each map polygon. The TCSBC approach is becoming standard in Canada and an ontology for surficial geology and surficial processes is integral to the TCSBC. Soil slides are detachment of regolith from underlying bedrock and travel as a block or flow-like mass across regolith or bedrock on slopes below (TCSBC units: R-bedrock; M-morainal sediments, primarily till; C-colluvial sediments; V-active gulleying; A-subject to snow avalanches). The components for a general model concerning soil slides are presented below.

GENERAL SOIL SLIDE COMPONENTS

Terrain units:

<u>Primary Terrain</u> unit is USUALLY M but SOMETIMES R
Commentary: In TCSBC, R can include minor areas of surficial deposits
<u>Secondary Primary Terrain unit</u> is USUALLY C if primary is R
<u>Minor terrain unit</u> will ALWAYS be M or C if _major terrain unit is R alone

Modifying processes:

SOMETIMES associated with V or A

Slope:

NEVER on slopes 14 degrees or less
SOMETIMES on slopes between 15 and 19 degrees
USUALLY on slopes between 20 and 40 degrees
RARELY on slopes 41 to 60 degrees
NEVER on slopes 60 to 90 degrees

The tendency operators ALWAYS, USUALLY, SOMETIMES, RARELY, and NEVER are based on the presence or absence of the variable in examination of factors associated with extant landslide environments in the area of failure. For our purposes, we used ALWAYS=100%, USUALLY 51-99%, SOMETIMES 20-50%, rarely 1-20% and NEVER=0%.

5. PRELIMINARY TRIAL

At this writing, a preliminary test of the system was made for portions of a 128,000 hectare region of the coast range in southwestern British Columbia covering 1:20,000 map sheets 92G.085, 92G.075, 92G.065, 92G.064, 092G.054, 092G.044, 092G.034, and part of 092G.034 (Fig. 3). Topography is mountainous and most areas up to approximately 1600 m are forested and have been extensively logged over recent decades. Although many types of landslides occur, post logging soil slides are the most numerous. These, along with other landside types were mapped and locations of failure crowns were digitized. Slopes and surficial geology were recorded in order to create a landslide occurrence model. Using conventional GIS techniques, the following data layers were spatially joined and clipped to map sheet boundaries:

1. Slope - derived from 25m cell slopes which was derived from the 25m cell digital elevation model
2. Bedrock geology
3. Geological faults and contacts
4. Surficial geology (BC terrain inventory legend)
5. Water features (rivers and coastline)
6. Transportation features (railway lines and roads)

This processing yielded between 2,000 and 20,000 polygons per map sheet. The polygon attributes were then translated to a standard science language with reference to the project ontology. This project ontology was prototyped and it included internationally recognised terminological standards, where available, such as those published by the British Geological Survey (1999) for rock nomenclature and Howes and Kenk (1997) for terrain classification. Once translation was complete, the attributes of each polygon were transferred into semantic networks (hierarchical data structures which respect the relationships, attributes and values of the original tabular data, but represent it in a format much more amenable to logical reasoning). The semantic networks and their associated maps were uploaded to the Internet environment making them available for assessment against (comparison with) any landslide susceptibility models present in the system, or which might be entered into it in the future. Figure 2 shows an example output with only the three highest similarity score intervals portrayed. The original map was in colour but it was converted to black and white for this publication. The system currently uses the addition of log probabilities to calculate two similarity scores for each comparison made: the first being a 'raw score' whose maximum possible value increases as a function of the number of attributes specified in the model description. The second is a score normalized

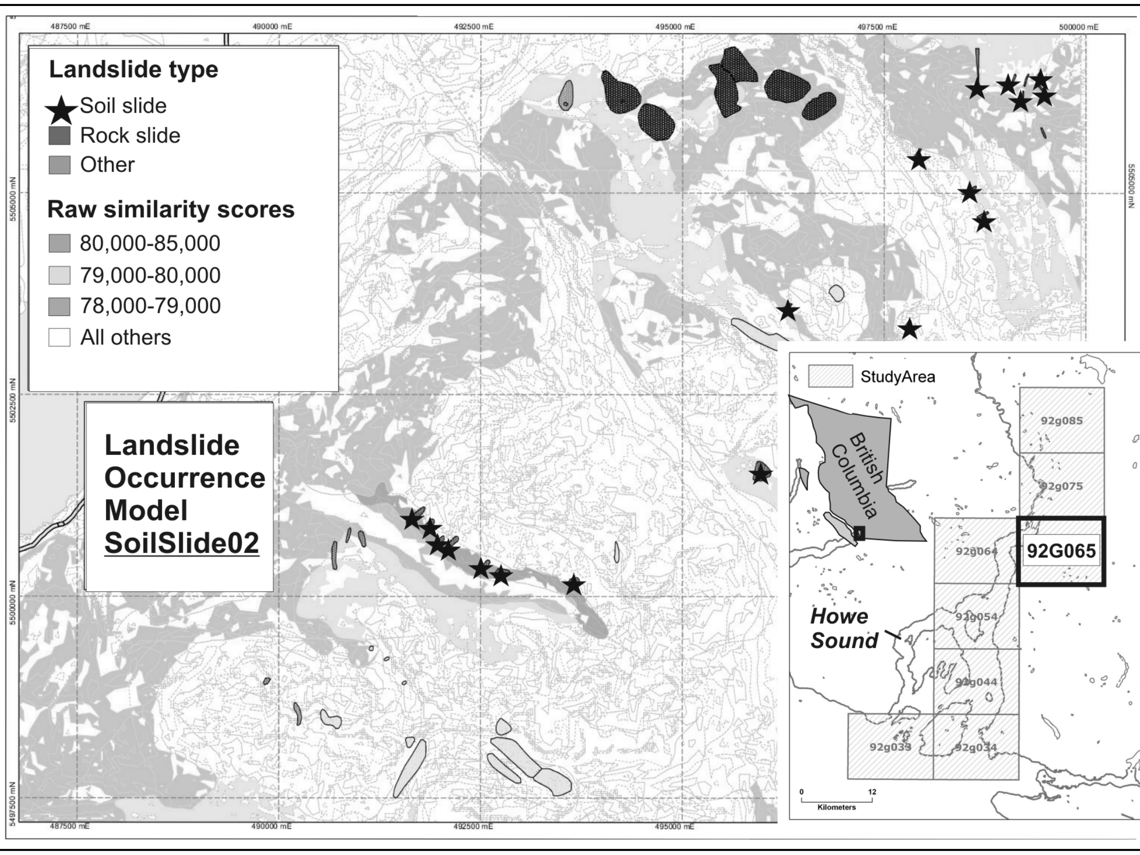

Figure 2. Map output displaying raw similarity scores from soil slide model SoilSlide02—one of several variations used in the preliminary trial.

to the range 0 to 100, such that the maximum possible (for a perfect match) is 100, and the worst possible match is (approximately) 0. The normalized score is useful for preliminary comparison studies, as it provides the user with a quick intuitive feeling for the closeness of match between two compared descriptions. To reach the same understanding of a comparison scored only with a raw score, the user would first need to know the raw score expected from a perfect match. More in-depth studies, however, are best undertaken using the raw scores. Work is ongoing in developing a new more mathematically robust method for similarity score calculation.

The results of this preliminary trial appear very promising: in the case of the landslides mapped in 92G065, 19 post-logging soil slides are present (failures along logging roads were not included). 17 out of 19 of the soil slides fall within polygons with similarity scores >78,000. Using that score as a cut off, 89% of the soil slides in the area are accounted for by the occurrence model. If this output were to be used as a susceptibility map for post-logging soil slides, scores above 80 000 could be qualitatively assigned as high and 78 000 to 80 000 could be assigned as moderate based upon relative soil slide density. A more rigorous and consistent method for designating rating values based upon raw scores will be addressed during further development of this technology.

7. ONGOING DEVELOPMENT ACTIVITIES

Several limitations were identified in this development phase of the system. These are currently being addressed.

7.1 Slope aspect

The current system cannot accept slope aspect in degrees as an input. Slope aspect is an important variable (e.g. Sawyer and Butler, 2004) for a variety of reasons including differences in microclimate (affects moisture contents of slopes and weathering processes) and structural and stratigraphic controls on slope formation and strength of earth materials as determined by structural history and lithology. An ability to reason about slope aspect across the numerical discontinuity between 359 and 1 degrees – specifically about the extent to which different slope aspects are similar to each other will be included in the next version of the systems' ontology. Reasoning with aspects measured in degrees on the compass is non-trivial. There are three contexts in which this difficulty can present itself, each requiring a different solution:

1. When aspects compared are expressed as single directions (e.g. 045 degrees);
2. When aspects compared are expressed as ranges (e.g.: 350 to 010 degrees);
3. When one of the aspects to be compared is expressed as a range, and the other is expressed as a single direction.

Part of the solution to this problem lies in using 'interval reasoning' on the difference between compared aspects, as calculated by subtracting the larger aspect (azimuth) from the smaller aspect, and then subtracting the result from 360 if the result is greater than 180. Decisions will need to be taken as to how to score aspect similarity as a function of difference in aspect. For example, will an aspect difference of up to ten degrees be considered to be a 'perfect' match, scoring maximum similarity points? The context for making such decisions has been set in Huang *et al.* (2005).

7.2 User-controlled weighting of attribute importance

Users of the system have expressed a desire to manipulate the 'importance weighting' accorded to attributes during the calculation of similarity scores. This need has been expressed in spite of the different weightings already given by the system to attributes based on whether they are ALWAYS, USUALLY, SOMETIMES, RARELY or NEVER expected by the model. Space is insufficient here to deal with this difficult problem that bears upon several fields of computer science and statistics (Sharma *et al.* 2007). Additional weighting options are currently in the final stages of development.

7.3 Comparison with multivariate methods

A more extensive application of the system is anticipated in 2008 within the area of the 'Sea-to-Sky Corridor' that links Vancouver with Whistler based upon recent digital surficial geology and landslide mapping. This area contains many types of landslides from rock falls and avalanches to soil slide rotational slumps in bedrock. Rock types include crystalline rocks of the Coast Plutonic Complex, metavolcanic and metasedimentary rocks and young volcanics. Landslide susceptibility mapping using conventional GIS-based multivariate analysis will be generated for the same area. The output from of our AI–based system will be compared to that of the multivariate methodology so that its strengths and weaknesses can be further evaluated.

9. ACKNOWLEDGMENTS

The authors are indebted to Andrée-Blais-Stevens, Peter Bobrowsky, Murray Journeay and Courtney Jermyn for critical reviews of this paper. However, all errors or omissions are purely the responsibility of the authors.

!0. REFERENCES

AllegroGraph 2.0 Triples Store 2007 http://www.franz.com/products/allegrograph/

British Geological Survey 1999. BGS Rock Classification Scheme, Vols. 1 – 4,http://www.bgs.ac.uk/bgsrcs

British Columbia Ministry of Environment, Lands & Parks 1997. Terrain Classification System for British Columbia Version 2. http://srmwww.gov.bc.ca/risc/pubs/teecolo/terclass/indes.html

Cruden, D.M and Varnes, D.J. 1996, Landslide types and processes In Turner and Schuster (eds.) Landslides—investigation and mitigation. Transportation Research Board Special Report 247, p. 36-75, National Academic Press, Washington, D.C.

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**: 181-216.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., Gali, M. (2006). "Estimating the quality of landslide susceptibility models." Geomorphology **81**: 166-184.

Howes, D. and Kenk, E. 1997. Terrain classification system for British Columbia, Version 2 (http://ilmbwww.gov.bc.ca/risc/pubs/teecolo/terclass/cove1.htm)

Huang, E., Smyth, C. and Poole, D. 2005. Representing and reasoning with Intervals. *2005 Annual Conference of the International Association for Mathematical Geology* http://www.georeferenceonline.com/ref/iamg_paperHuangSmythPoole2005.pdf

McGuiness, D.L., and van Harmelen, F., 2004. OWL Web Ontology Language Overview.http://www.w3.org/TR/owl-features/

OneGeology Initiative 2007. http://www.onegeology.org/

Roddick, J A; Woodsworth, G J; Hutchinson, W W, 1979.Geology of Vancouver west half and Mainland part of Alberni, British Columbia. Geological Survey of Canada Open File 611.

Sawyer, Carol F., and David R. Butler 2004. Landslide aspect: a methodological approach to circular data for hazard analysis. *Papers of the Applied Geography Conference 27, St. Louis, MO, 67-74.*

Sharma, R., Poole, D. and Smyth, C. 2007. A System for ontologically-grounded probabilistic Matching. The 5th Bayesian Modeling Applications Workshop, UAI, Vancouver, British Columbia

Thompson, B. 1980. Terrain Inventory, Squamish. British Columbia Ministry of Energy, Mines and Petroleum Resources, 1:50 000

TopBraid Composer Ontology Editor 2006. http://www.topbraidcomposer.com/

Suzen, M. L., Doyuran, V. 2006. A comparison of the GIS based landslide susceptibility assessment methods: multivariate vs. bivariate. Environmental gelology **45**: 665-679.

Wang, H. B., Sassa, K., Xu, W.Y. (2007). "Assessment of landslide susceptibility using multivariate logistic regression: a case study in southern Japan. Environmental and engineering geoscience XIII(2): 183-192.