
ArcGIS V.10 Landslide Susceptibility Data Mining Add-in Tool Integrating Data Mining and GIS Techniques to Model Landslide Susceptibility

208

Darshika Palamakumbure, David Stirling, Phil Flentje, and Robin Chowdhury

Abstract

Landslide susceptibility modeling is an essential early step towards managing landslide risk. A minimum of \$4.8 million is lost due to landslide related damages every year in Illawara region of Australia. At present, Data mining and knowledge discovery techniques are becoming popular in building landslide susceptibility models due to their enhanced predictive performances. Until now, the lack of tools to undertake data extraction and making the predictions have limited the applicability of this novel technique in landslide model building. This paper discusses the development of the LSDM (Landslide Susceptibility Data Mining) toolbar which was designed to utilize machine learning techniques within a GIS environment by coupling GIS and data mining software (See5) capabilities. The software development kit available with ArcGIS v.10 has been utilized in developing the toolbar add-in. The fundamental tasks; data preparation, model optimizing, derivation of decision trees, predictions and validation are all performed using the individual controls available in the toolbar. This tool automates the entire model building process and in preparation of training data and producing outcomes that are compliant with both national and international Landslide Risk management guidelines.

Keywords

Data mining • GIS • Landslide inventory • Landslide susceptibility • See5

208.1 Introduction

Landslides are one of the most catastrophic geo-hazards that cause severe injuries, property destruction and deaths all over the world. The challenge that landslides have posed to property and infrastructure and human lives over the past few decades in Australia, have been summarized by Flentje (1998) and, Leventhal and Kotze (2008). In the Wollongong local government area in NSW, landslide costs have been

reported by Osuchowski and Roberts (2011), and Tobin (2012). The total minimum estimated value of destroyed and damaged houses due to landsliding in Wollongong region is AUD \$14 million and the cost of nearly AUD \$300 million has been borne by government authorities over the period of 1950–2013 to restore landslide-related damages. This equates to a minimum annual expenditure of AUD \$4.8 million.

With the increasing pressure to develop marginal land, consideration of landslide hazard and risk in land-use planning has become more important in recent decades. In order to assess and manage the landslide risk, the Australian Geomechanics Society Landslide Risk Management Guidelines (AGS 2007) and JTC 1 2008 (Fell et al. 2008) guidelines logically suggest that development of Landslide Inventories and then Landslide Susceptibility Zoning are essential. Application of decision trees (a knowledge-based data-mining technique) (Kohavi and Quinlan 1999; Quinlan

D. Palamakumbure (✉) · D. Stirling · P. Flentje (✉) · R. Chowdhury
University of Wollongong, Northfields Avenue, Wollongong NSW 2522, Australia
e-mail: dp770@uowmail.edu.au

P. Flentje
e-mail: pflentje@uow.edu.au

1986) in landslide susceptibility mapping has been gaining attention over past few decades due to its enhanced predictive capabilities and independence from subjective expert judgments (Flentje et al. 2007; Miner et al. 2010; Saito et al. 2009; Yeon et al. 2010). Until now, there have been limitations in integrating data mining and GIS techniques due to incompatibilities between data mining software requirements and readily accessible GIS datasets and tools. Besides converting GIS data into a structure which is readable in the data mining software See5 (Quinlan 1993), returning the See5 output and interpreting the confidence grid within a GIS environment is another important challenge to overcome when using data mining techniques for modelling of landslide susceptibility. The University of Wollongong Landslide Research Team has developed the LSDM toolbar for these reasons. These tools extract and convert data from any GIS layers (including from the Landslide Inventory), format the data to meet the input requirements of the See5 Data Mining software, call the See5 program, apply the results of the decision tree analysis to produce a validated numerical grid of landslide susceptibility and classify it according to the recommendations of the Landslide Risk Management guidelines (AGS 2007; Table 4).

208.2 Implementation of LSDM ArcGIS Add-in Toolbar

ArcGIS v.10 is enhanced with the embedded scripting language Python and a new desktop customization VB.NET add-In module. The customization capability of the ArcObjects software development kit integrated with .NET development environment (Visual Studio Express for .NET 2008), has been used in developing the LSDM toolbar. Working with ArcObjects (a library of Component Object Model components which forms the base of the ArcGIS) within a .NET development environment, enables access to a series of ArcGIS built-in tools such as data management, visualization and spatial algorithms. This has allowed the writers to develop the LSDM toolbar relatively quickly.

The Visual studio.net environment is used to implement the interface for receiving user commands. The LSDM toolbar comprised of six command buttons (Fig. 208.1) to conduct data mining and GIS tasks entirely within the desktop GIS environment.

The landslide inventory (Flentje et al. 2012) is the most important evidence based input layer that is used in the preparation of landslide susceptibility and hazard zoning. A DEM and its derivatives and geology are the next primary inputs for building the model, but any layers deemed

appropriate can be used. See5 gains information pixel by pixel from each data layer corresponding to landslide and non-landslide locations, the first tool component (button) in the LSDM toolbar records pixel X, Y location with the corresponding attribute values from each of the input model layers. DEM raster layer and other derivatives can be selected by the user to extract cell values into a text file with the extension of .data which is readable in See5 software. Also, a *.names file is created based on the user inputs to describe the attributes of the corresponding data layers. The second component of the tool calls See5 (a standalone license for the low cost Data Mining software See5 must be installed on the host workstation) to derive the rule sets from the input data. Analyzing the rules and calibrating the susceptibility model is also done automatically at this stage. The third component of the LSDM toolbar is used to re-map See5 rules into GIS map algebra functions. These functions apply the logic of the rules using the input data layers so that the modelled outcome merges all the See5 rule based predictions into a new floating point ESRI grid. This grid represents the landslide susceptibility with a numerical value assigned to each cell location. Also, this component provides the facility to produce the source files to draw Receiver Operating Characteristic (ROC) curves (Corominas et al. 2013) and success rate curves to assess the accuracy of the susceptibility model with respect to the Landslide Inventory distribution.

The fourth and final component can be used to visualize the ROC curves and success rate curves. Furthermore, this feature facilitates the assessment of susceptibility class boundaries (user defines the number of classes). These can be obtained by entering the percentage distribution of the landslide inventory being included in a particular class and based on these parameters; a classified map layer of the landslide susceptibility map is produced.

208.3 Application

A revised landslide susceptibility model has been developed using the LSDM toolbar for the Wollongong local government area in order to obtain an accurate and an up-to-date outcome using a high resolution DEM derived from the latest airborne laser scan data. Before arriving at the final result, several iterations of the susceptibility model building can be required to select the most significant landslide causative factors and filter off datasets or individual parameters which may mislead the model. A total number of 5,89,060 landslide and non-landslide pixels have been recorded with the attributes of 10 landslide causative factors identified as most

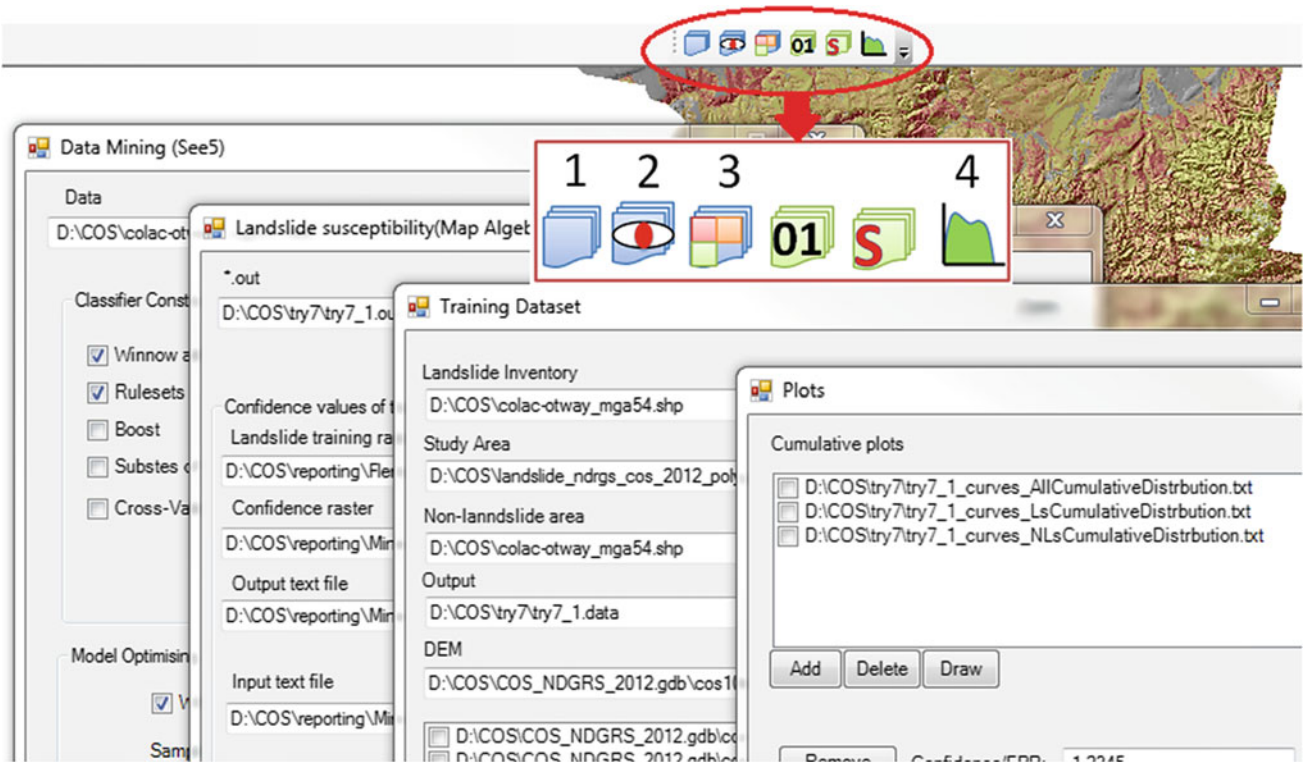
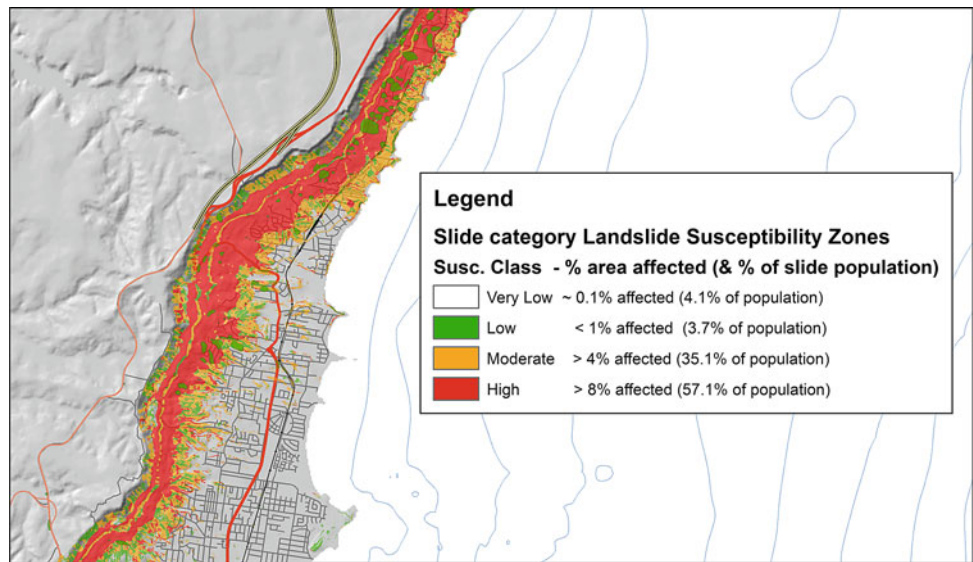


Fig. 208.1 LSDM toolbar

Fig. 208.2 Landslide susceptibility map of Wollongong local government area



relevant in the latest Wollongong study. A Landslide susceptibility model has been developed using 211 rules which depict the classification process of landslides and non-landslide pixels according to the cut-off values of one or more data layers. In the resulting susceptibility map, the high susceptibility class includes 57.1 % of the total landslide population whereas very low susceptibility class includes 4.1 % (Fig. 208.2).

208.4 Conclusion

The LSDM tool has been successfully trialled on very large grids of 3 billion pixels and it has been proven to be a powerful tool in providing resilient, quantifiable and repeatable landslide susceptibility models. The customization capabilities available with Arc-GIS have enabled the automation of the GIS data

preparation for data mining, the actual data mining, calling the See5 software from within Arc and converting the final outcome into an ArcGIS grid. The LSDM toolbar rigorously classifies this grid into susceptibility classes based on the landslide distribution as per the AGS (2007) guidelines. This development has been very successful in saving time and providing a user friendly interface with built-in grid classification tools to produce an accurate and transparent outcome, essentially free from subjective expert user judgments. The integration of See5 and GIS techniques enables visualizing of the final rule based modelling outcome and assessing the accuracy both qualitatively and quantitatively. The alternative to date has been a highly time consuming and tedious series of manual processes. Given a landslide inventory and relevant data layers, months of work are now reduced potentially to a few minutes or hours of simple processing time. Upon completion of the first authors PhD, expected in late 2014, this LSDM toolbar will be made available at no cost from our University of Wollongong landslide research team website (<http://eis.uow.edu.au/landslide>).

References

- AGS (2007) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Aust Geomech J* 42(1):23
- Corominas J, Westen C, Frattini P, Cascini L, Malet JP, Fotopoulou S, Catani F, Eeckhaut M, Mavrouli O, Agliardi F, Pitilakis K, Winter MG, Pastor M, Ferlisi S, Tofani V, Hervás J, Smith JT (2013) Recommendations for the quantitative analysis of landslide risk. *Bull Eng Geol Environ* 1–55. doi:10.1007/s10064-013-0538-8
- Fell R, Corominas J, Bonnard C, Cascini L, Leroi E, Savage WZ (2008) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Eng Geol* 102(3–4):85–98. doi:10.1016/j.enggeo.2008.03.022
- Flentje P (1998) Computer based landslide hazard and risk assessment: Northern Illawarra region of New South Wales, Australia. Ph.D., University of Wollongong New South Wales, Australia. <http://eis.uow.edu.au/landslide/flentjephthesis/index.html>
- Flentje P, Stirling D, Chowdhury R (2007) Landslide susceptibility and hazard derived from a landslide inventory using data mining: an Australian case study. Proceedings of the first North American landslide conference, landslide and society: integrated science, engineering, management and mitigation, Vail, Colorado, 3rd–8th June 2007. CD, Paper number 17823–024
- Flentje P, Stirling D, Palamakumbure D (2012) An inventory of landslides within the Sydney Basin to aid the development of a refined susceptibility zoning. Proceedings of the 11th Australia New Zealand conference on geomechanics (ANZ 2012), ground engineering in a changing world, Melbourne, Australia, 15th–18th July 2012
- Kohavi R, Quinlan R (1999). Decision tree discovery. Handbook of data mining and knowledge discovery. Oxford University Press, Oxford, pp 267–276
- Leventhal AR, Kotze GP (2008) Landslide susceptibility and hazard mapping in Australia for land-use planning: with reference to challenges in metropolitan suburbia. *Eng Geol* 102(3–4):238–250. doi:10.1016/j.enggeo.2008.03.021
- Miner AS, Vamplew P, Windle DJ, Flentje P, Warner P (2010) A comparative study of various data mining techniques as applied to the modeling of landslide susceptibility on the Bellarine Peninsula, Victoria, Australia. Proceedings of the 11th IAEG congress of the international association of engineering geology and the environment, Auckland, New Zealand
- Osuchowski M, Roberts J (2011) Landslide costs in the Wollongong Region. Monica Osuchowski and Jenna Roberts: Geoscience Australia. ISBN: 9781921954337
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106. doi:10.1023/a:1022643204877
- Quinlan JR (1993) C4.5: programs for machine learning: Morgan Kaufmann Publishers Inc., San Francisco. ISBN: 1-55860-238-0
- Saito H, Nakayama D, 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:10.1016/j.geomorph.2009.02.026
- Tobin PR (2012) Manage slope instability hazards affecting local roads within the city of Wollongong. *Aust Geomech J* 47:53–58
- Yeon YK, Han JG, Ryu KH (2010) Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Eng Geol* 116(3–4):274–283