The paper was published in the proceedings of the 12th Australia New Zealand Conference on Geomechanics and was edited by Graham Ramsey. The conference was held in Wellington, New Zealand, 22-25 February 2015.

Flow category landslide susceptibility modelling of the Sydney Basin

D. Palamakumbure, P. Flentje and D. Stirling

[1]Faculty of Engineering and Information Sciences, University of Wollongong, NSW 2522; email: pflentje@uow.edu.au

ABSTRACT

The University of Wollongong Landslide Research Team has completed a GIS-based Landslide Susceptibility model for the entire Sydney Basin region. According to the Australian Bureau of Statistics and the 2011 Census data, the population within the Sydney Basin Study area is approximately one quarter of the population of Australia. This model has been developed with the aid of a large scale Landslide Inventory for NSW, which contains 1823 landslides to date. A composite geology dataset has also been developed using commercially available geology datasets including those from NSW Department of Primary Industries and elsewhere. The model employs a 10m pixel Digital Elevation Model (DEM) across the entire study area derived from either Local Government sourced Airborne Laser Scan data and where absent the 30m pixel year 2000 Shuttle Radar Topography Mission (SRTM) data. Using techniques developed over the last decade and refined ArcGIS tools developed over the last three years, Data Mining methods and ESRI ArcGIS capabilities have enabled the modelling to produce a very useful zoning outcome over the entire Sydney Basin area. The Major advantage of this new tool is that it applies the See5 logic derived from rule sets over a large datasets, and produces a visually interpretable outcome. The authors expect the susceptibility zoning are suitable for use at Regional to Local Advisory level Local Government Planning Development Control Plans.

Keywords: landslides, susceptibility modelling, GIS, See5, flows

1 INTRODUCTION

This paper discusses the progress of flow category landslide susceptibility modelling of the Sydney Basin study area. After compiling of major datasets for the entire Sydney Basin study area, a susceptibility model for flows was developed along with the slide category landslide susceptibility modelling. The Sydney Basin study area region extends from Newcastle in the north to Batemans Bay in the south and west to include the Blue Mountains, an area of 30,603 km^{2}in NSW, Australia. The Australian Bureau of Statistics and the 2011 census data reports that the population within this area is 5.4 million people, approximately one quarter of the population of Australia. Therefore, proper land-use planning is considered essential to cope up with the increasing pressure to develop marginal land. In local government areas where catastrophic landslides have occurred, Landslide Risk Assessment and management, is recognised as important for proper land-use zoning practices. The Landslide Risk Management Guidelines (AGS, 2007) and JTC-1 2008 (Fell et al., 2008; Fell et al., 2008) state the development of Landslide Inventories and then Landslide Susceptibility Zoning as the first step of landslide risk assessment for effective land use planning.

Based on detailed and comprehensive landslide mapping, as a companion paper to slide modelling (Palamakumbure et al., 2014), this study focuses on using a data mining technique, namely decision tree derived rule-sets, to model the susceptibility of flow category landslides. In the literature, decision trees have been used to map landslide susceptibility in numerous occasions and this technique is well known for its enhanced predictive capabilities, transparency and interpretability (Flentje et al., 2007; Saito et al., 2009; Miner et al., 2010; Wang and Niu, 2010; Yeon et al., 2010). See5 data mining software (Quinlan, 2013) developed based on C5 learning algorithm, was used in this study to develop decision tree derived rules.

Expansion of the University of Wollongong (UOW) landslide inventory from its Illawarra centric coverage to include the landslides across the entire Sydney Basin study area has been undertaken by the Landslide Research Team (LRT) (Flentje et al., 2012). To November 2014, the inventory contains 1823 landslides, out of which 267 are flows. Figure 1 summarises the volume distribution of 93 flows, of which the detailed information is available in our inventory. Compilation of a high/medium resolution composite Digital elevation model and Geology datasets has been completed. With the data collection now being finalized, two susceptibility maps for slide and flow category landslides have been prepared. These susceptibility zoning outcomes are suitable for use as Preliminary or up to Intermediate level Susceptibility Zoning for Local Government Planning Development Control Plans in the absence of any other information.

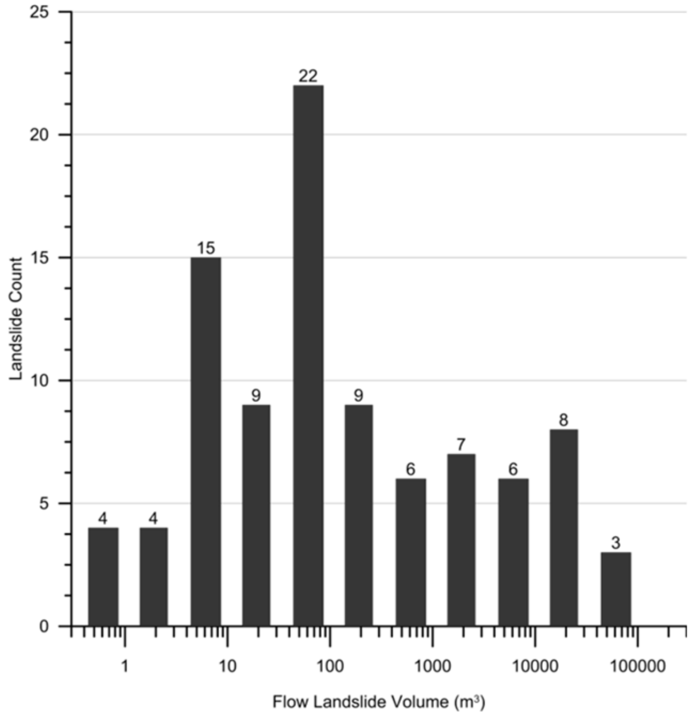

Figure 1. Volume distribution of 93 flows

2 DATA SETS AND TOOLS

2.1 Digital Elevation model

High resolution Airborne Laser Scan data (ALS) is available for some parts of the study area. In order to cover remaining parts of the study area, CSIRO/Geoscience Australia/NASA Global DEM V2.0 (NASA, 2011) at 30m was used. The high density ALS point cloud is suitable for preparing a high resolution DEM at 10m. Therefore, NASA GDEM was resampled into 10m grid cell size before combining it with the ALS DEM to produce a composite digital elevation model to cover the study area. Subsequently, from this DEM, eight other derivatives namely, *Slope, Aspect, Curvature, Profile Curvature, Plan Curvature, Flow Accumulation, Wetness Index* and *Terrain classification* were obtained as model input layers

2.2 Software tools

For the model development and multilayer data analysis, ArcGIS v.10 software environment was used. Furthermore, See5 software was used to derive decision tree based rule-sets. The entire data mining and GIS process was automated by developing an ArcGIS Landslide data mining (LSDM) add-in toolbar(Palamakumbure et al., 2014). This tool automates a series of tedious manual processes involved in data extraction, preparation, deriving See5 rules and preparation of the ArcGIS susceptibility grid.

3 LANDSLIDE PREDICTIONS AND THE SUSCEPTIBILITY

The ArcGIS LSDM toolbar has been used to finalise the process of extracting attributes of the GIS data layers, calling See5, applying rule based predictions over the study area and making the final susceptibility map. The training dataset was prepared by selecting all of the flow category landslide pixels and an equal number of non-flow pixels to balance the numerical output of the model. The attribute values of each input layers corresponding to all of the flow pixel locations and selected non-flow locations were extracted as separate training cases. The See5 constructs decision tree classifiers by defining test conditions based on the attribute values and splitting the training data into smaller subsets.

Normally the See5 learning algorithm being a discrete or categorical classifier predicts a discrete class corresponding to a case. However, according to the Landslide Risk Management (LRM) guidelines, landslide susceptibility has to be expressed as a continuous number. Therefore, real valued likelihood values were produced using confidence values of rules. Confidence of the predications made is evaluated using the Laplace ratio *(n-m+1)/(n+2)* where *n* is the number of training cases that a specific rule covers and *m*, is the number of wrongly classified cases. The average confidence value of the rules participated in classifying a pixel ranges from 0 to 1. When a pixel satisfies the conditions of landslide and non-landslide class rules, the class which holds the highest average confidence value wins. If the average confidence value of the non-landslide class is greater than that of the landslide class, the confidence of the non-landslide class prediction is given by multiplying the average confidence by -1. This method allows the landslide susceptibility to be presented with a value which ranges from -1 to 1.

4 ANALYSIS OF LANDSLIDE SUSCEPTIBILITY ZONES

Data from eight different layers derived from the Digital Elevation model was extracted corresponding to the landslide and randomly selected non-landslide pixel locations. Modelling of Slide category (Palamakumbure et al., 2014) and Flow category landslides have been conducted separately using the same See5 methodology and Table 1 summarises the results.

Table 1: Summary of Flow and Slide category landslide susceptibility modelling

	Flow	Slide
Attribute	Usage (%)	
Slope	100	38
Plan Curvature	39	7
Profile Curvature	26	11
Curvature	26	9
Aspect	16	11
Terrain	14	4
Wetness Index	12	12
Geology	-	100
Flow Accumulation	<1%	<1%
Training cases	32,862	670,164

Geology data layer was not used in modelling of the flow category landslides as it is assumed that the occurrence of flows does not largely depend on *Geology*. Debris flows are generally shallow seated landslides and therefore, underlying geology is less relevant. Our modelling methodology uses known debris flows as model training reference points. Our model is based on the mapped location of only 267 flows within an area of 30,603 km^2. If the Geology was included in the modelling, the spatial extent of the modelled Debris Flow susceptibility would be more limited by the Geology in which they occur, which we consider to be unnecessarily restrictive for the application – developing a debris flow susceptibility map with wide application. If alternatively, say 1000 debris flows had been mapped within a single local government area, modelling within that small area may likely be best done using geology.

For Slide category landslides, as shown in Table 1, *Geology* has contributed to classify 100% of the data and the second largest amount of data was classified using *Slope*. When modelling of flows, *Slope* has classified 100% of the data. *Plan Curvature, Profile Curvature, Curvature* and *Terrain classification* have classified more data in modelling of flows than that of the slides and the contribution of *Flow accumulation* was negligible in both models.

The values in the table 4(b) of AGS LRM Zoning Guidelines (AGS, 2007) have been used as a reference to categorising the landslide susceptibility classes relative to the landslide inventory. As our inventory is quite accurate, albeit incomplete, we regard table 4(b) as being most appropriate. The logic of the See5 rules has been applied across the entire raster grids producing a landslide confidence value for each pixel. Figure 2 plots the landslide confidence value against the cumulative percentage of pixels for 1) all the mapped flow category landslides and 2) the entire model area. Following the steps in the distribution curves and the requiems of Table 4(b), four landslide susceptibility classes were defined as per the demarcated four regions in the Figure 2. Flow pixels curve and study area pixels curve in each region aid the calculation of study area and landslide area in each susceptibility class. Figure 2 clearly shows that 50% of our inventory is captured in just 15% of the study area, and furthermore, 80% of our inventory is captured in 28% of the study area further reflecting that table 4(b) is best adopted with our work. It is also of note that the second author of this paper helped develop this particular portion of the AGS Guidelines during early iterations of the work reported herein.

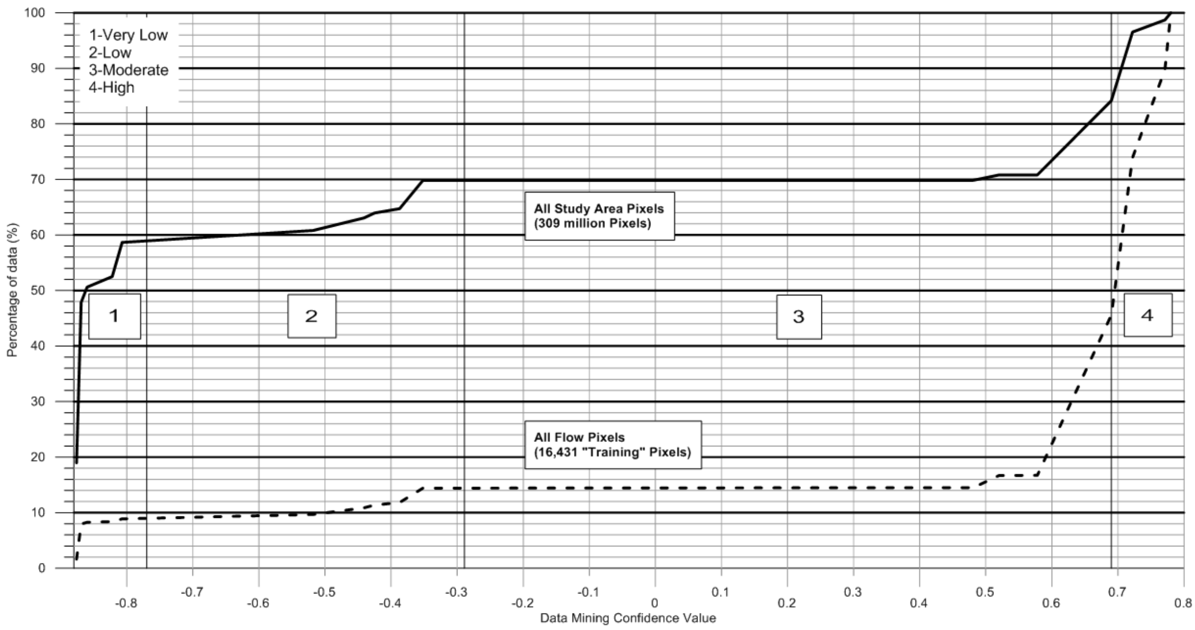

Figure 2. Classification of Susceptibility Zones using the distribution of the confidence values

The Susceptibility modelling of Flow category landslides (Table 2) has classified 16% of the study area (approximately 4,944 km^2), as High Susceptibility. This area contains 54% of the known flows with a density of 0.02%. The moderate susceptibility class covers nearly 14% of the study area (4,326 km^2) and contains 32% of the flow population with a flow density of 0.01%. The area of Low Susceptibility zone is 3,399 km^2 (11% of the study area) and contains 5% of the flow population with a flow density of 0.002%. Almost 59% of the study area, approximately 18,233 km^2, has been classified as Very Low Susceptibility containing 9% of the flow population with a density of 0.0008%. Furthermore, considering the combined results of High and Moderate susceptibility classes, nearly 86% of the slides occur in just 30% of the study area.

Table 2: Distribution of flows in the landslide susceptibility classes.

Susceptibility Class	% of the Study Area	Area (km^2) of class	% of Flow population	Area of Flows (km^2)	% of zoned area effected by flows
Very Low - 1	59	18,233	9	0.15	0.0008
Low - 2	11	3,399	5	0.08	0.0024
Moderate - 3	14	4,326	32	0.53	0.0122
High - 4	16	4,944	54	0.89	0.0179

The percentage of landslides included in the Very Low category of the flow model (Table 3) is greater than that of the slide model and 8% higher than the recommended value in the Table 4(b) of LRM Guidelines (AGS, 2007). Furthermore, the High susceptibility class of the flow model covers 16% of the study area whereas in the slide model, the corresponding value is 6.5%. The area of the Very Low class of the flow model is 10% greater than that of the slide model. The number of training points available to train the slide category susceptibility model is almost 20 times greater than that of the flow category susceptibility modelling. Furthermore, the proportion of the each susceptibility class affected by flow category landslides is lower than the corresponding values of the slide category model outcome. The flow category landslide susceptibility map is shown in the Figure 3.

Table 3: Comparison of the susceptibility descriptors of Flow and Slide category models

Susceptibility Descriptors	Recommended % of landslides as in Table 4(b) of LRM Guidelines (AGS 2007)	% landslides		% study area		% zoned area effected	
		flows	slides	flows	slides	flows	slides
Very Low - 1	0 to 1	9	0.4	59	69.6	~0	~0
Low - 2	>1 to 10	5	3.5	11	15.5	0.002	0.19
Moderate - 3	>10 to 50	32	15.7	14	8.4	0.01	0.02
High - 4	>50	54	80.4	16	6.5	0.02	1.32

5 FIELD CALIBRATION OF THE FLOW MODEL

During the field data collection over a period of many years, a total of 503 field based assessments of flow category Landslide Susceptibility were recorded to facilitate subsequent model calibration. The field assessment work was undertaken by the first two authors and other colleges at different times. The work was completed using GPS/GNSS to record accurate spatial positioning, and assessing the susceptibility of an area equating to a 50m diameter circle (considered to be an appropriate area upon which to make a field judgement) centred at the recorded location. Numerical values of 1 to 4 were assigned to each of the field assessment locations from very low (189 points), low (174 points), moderate (95 points) to high (45 points) flow category landslide Susceptibility respectively. Using ESRI ArcGIS Spatial Analyst Zonal Statistics, the mean computer modelled Susceptibility value for all pixels within each of 50m diameter GIS-generated circles of approximately 1,963 square meters centred on each of the GPS recorded locations was determined. Then the modelled susceptibility was compared with the field based assessment.

The difference, D, between the average value predicted by the model (50m diameter circle, 1963.5m^2, intersecting all 10m pixels (100m^2)) and the value assessed independently in the field was plotted in the histogram shown in Figure 4. Therefore the difference D = 0 indicates the count for which the assessments match. Results are rounded to the nearest whole number. Almost 47% of the sites have average model results the same as they have been assessed in the field. An additional 16%, have been assessed by the computer model to be one Susceptibility class greater (the model is conservative) than that during the field assessment, and additional 4% has been assessed to be two Susceptibility classes greater than the field assessment. A further 22% have been assessed to be one Susceptibility class less than (the model is not conservative) that during the field assessment, with a further 9%, two classes less than the field assessments and 3%, three classes less than the field assessment.

6 CONCLUSION

The NSW Landslide Inventory and large scale GIS based data layers have been used in the modelling of the flow category Landslide Susceptibility. The See5 based data mining approach was successful in meeting the AGS (2007) Table 4(b) objectives up to a large extent. The slide category susceptibility model has been more successful in producing values that match the recommended susceptibility descriptors of the guidelines than the flow category model. This is due to the smaller number of flows (267), recorded in the Inventory relative to the number of slides (1424).

The Landslide Susceptibility toolbar has demonstrated its suitability for application in modelling large scale high resolution datasets. Based on the research work completed recently, the ratio between positive and negative training cases is chosen as 1:1. Also, further research work is still proceeding regarding selection of See5 modelling parameters suitable to conduct a large scale and high resolution modelling work. Assembling and preparation of data was one of the main challenges in this project and in particular the Landslide Inventory.

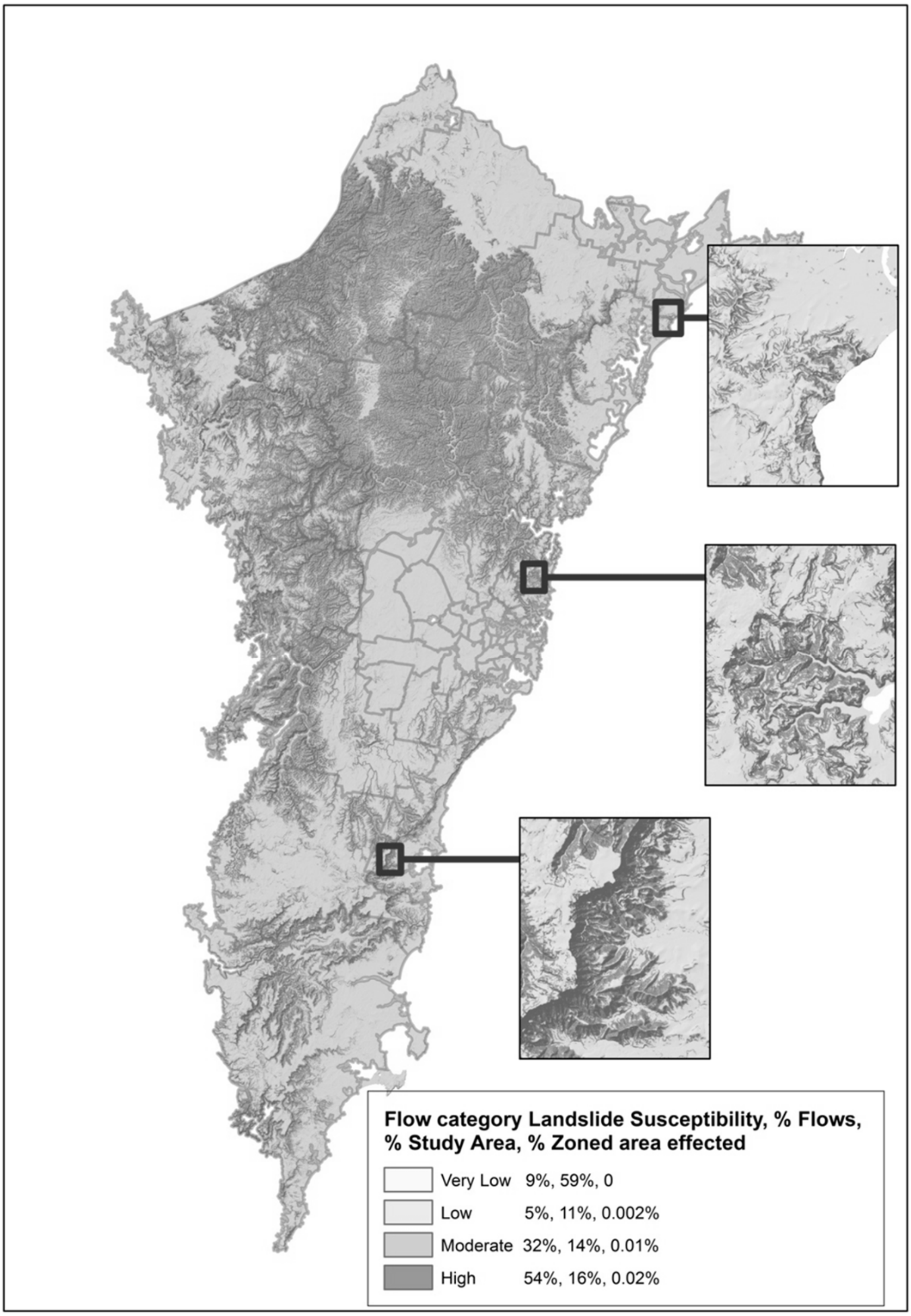

Figure 3. Flow category landslide susceptibility map of the Sydney Basin

This being a regional spatial model, rainfall intensity has not been incorporated in the modelling work as the data is hugely variable and extremely difficult to predict. Efforts have been made to include ground hydrogeology parameters as best as we can. It was noted that *Flow Accumulation* was the least contributing factor towards classifying data in both slide and flow models. *Geology* has not been considered as an important parameter in the regional flow modelling but when modelling slides, it was the main contributor towards classifying the data. In both models, *Slope* has been highlighted as an important parameter. Furthermore, in the flow category landslide susceptibility model, all of the *curvature* parameters have contributed more towards classifying the data than in the slide model. However, Wetness Index has been more useful in classifying slides than flows.

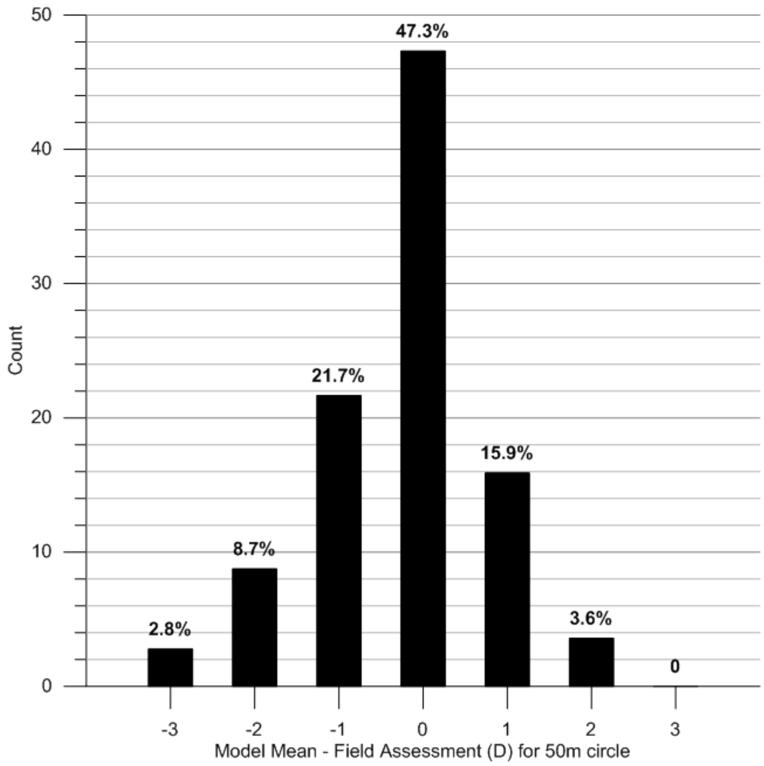

Figure 4. Histogram showing the difference between the Field and Modelled Landslide Susceptibility

We have compared the field assessment with model predictions and evaluated these comparisons. The results of the field assessment show that the model has an overall 67% of conservative success (D = 0, 1 and 2). The authors suggest that the flow Category Susceptibility Zoning outcomes may be suitable for use as Preliminary and perhaps up to Intermediate level Susceptibility Zoning for Local Government Planning Development Control Plans where no better zoning information exists. The modelling should differentiate between man-made and natural failures although we have not progressed to that level of work thus far. The inventory does differentiate man-made failures although more data regarding these types of failures does need to be collected. It is an area for future development.

The authors would like to provide this information to local governments in exchange for landslide inventory information. We would then be able to enhance our existing inventory, subject to funding and in turn iterate and further develop the modelling and zoning outcomes. We look forward to working with local governments across the Sydney Basin over the coming years.

7 REFERENCES

AGS (2007). "Guidelines for Landslide susceptibility, Hazard and Risk Zoning for Land Use Planning ", Australian Geomechanics Journal, vol. 42, issue (1), 23.

Fell R., Corominas J., Bonnard C., Cascini L., Leroi E. and Savage W. Z. (2008). "Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning." Engineering Geology, vol. 102, issue (3-4), 99-111.

Fell R., Corominas J., Bonnard C., Cascini L., Leroi E. and Savage W. Z. (2008). "Guidelines for landslide susceptibility, hazard and risk zoning for land use planning." Engineering Geology*,* vol. 102, issue (3-4), 85-98.

Flentje P., Stirling D. and Chowdhury R. (2007). Landslide susceptibility and Hazard derived from a landslide inventory using Data minig - an Australian case ctudy. First North American Landslide Conference, Landslide and Society: Integrated Science, Engineering, Management and Mitigation, Vail, Colarado.

Flentje P., Stirling D. and Palamakumbure D. (2012). An Inventory of Landslides within the Sydney Basin to aid the development of a refined Susceptibility Zoning. 11th Australia New Zealand Conference on Geomechanics (ANZ 2012), Ground Engineering in a Changing World, Melbourne, Australia, Australian Geomechanics

Miner A. S., Vamplew P., Windle D. J., Flentje P. and 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. 11th IAEG Congress of the International Association of Engineering Geology and the Environment, Auckland, New Zealand.

NASA (2011). ASTER GDEM v10. T. The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) is concurrently distributed from the Ministry of Economy, and Industry (METI) Earth Remote Sensing Data Analysis Center (ERSDAC) in Japan and the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Data Information System (EOSDIS) Land Processes (LP) Distributed Active Archive Center (DAAC) in the United States.

Palamakumbure D., Flentje P. and Stirling D. (2014). Landslide Inventory and Susceptibility Modelling of the Sydney Basin. 10th ANZ Young Geotechnical Engineering Conference (10YGPC), Noosa, Brisbane

Palamakumbure D., Stirling D. and Flentje P. (2014). ArcGIS v.10 Landslide Susceptibility Data Mining add-in tool integrating data mining and GIS techniques to model landslide susceptibility. IAEG 2014, Torino (In press), Springer.

Quinlan J. R. (2013). "Data Mining Tools See5 and C5.0." GNU General Public License, Release 2.07 GPL Edition. Retrieved 10/01/2012, 2012, from http://www.rulequest.com/.

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*,* vol. 109, issue (3-4), 108-121.

Wang X. and Niu R. (2010). "Landslide intelligent prediction using object-oriented method." Soil Dynamics and Earthquake Engineering*,* vol. 30, issue (12), 1478-1486.

Yeon Y. K., Han J. G. and Ryu K. H. (2010). "Landslide susceptibility mapping in Injae, Korea, using a decision tree." Engineering Geology*,* vol. 116, issue (3-4), 274-283.