Assessing the risk landslides pose to road and rail networks

Taalab K P[*1] and Cheng T[†1]

[1] SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, Chadwick Building, Gower Street, London WC1E 6BT, United Kingdom

November 5, 2014

Summary

Road and rail networks are critical infrastructure, vital for ensuring the flow of essential goods and services necessary to maintain a country's economic and national security. Landslides are a natural hazards which can seriously affect these networks, so in order to plan mitigation strategies, calculate losses and minimise casualties, it is necessary know the risk a posed by landslides. Using a case study of Piedmont, Italy, this study proposes an empirical modelling approach to the quantification of landslide risk using support vector machines and simple network analysis.

KEYWORDS: Landslide, Infrastructure, Support vector machine, Hazard, Risk

1. Introduction

Road and rail networks are critical infrastructure, vital for ensuring the flow of essential goods and services necessary to maintain a country's economic and national security. Landslides are a natural hazards which can seriously affect road and rail networks, so in order to plan mitigation strategies, calculate losses and minimise casualties, it is necessary know the risk a posed by landslides. The identification of risk can be deconstructed into a number of constituent parts. Varnes (1984) proposes that risk can be calculated using the formula:

$$Risk = (susceptibility \ x \ trigger) \ x \ (vulnerability \ x \ exposure) \quad (1)$$

where *susceptibility x trigger = hazard*. Hazard susceptibility is a relative measure of the spatial likelihood of the occurrence of landslides (Pourghasemi et al., 2013). It can be determined as a function of terrain attributes (e.g. slope, aspect) and environmental variables (e.g. geology, soil and land use). Susceptibility alone cannot give the probability of landslide occurrence. For this, it is necessary to model the relationship between events which trigger landslides and landslide occurrence. There are many potential triggers including precipitation, earthquakes and human activity, with heavy rainfall being the most common trigger (Cepeda et al., 2010). As an event in itself, a landslide does not pose any risk. The risk comes from the exposure of people or elements of the build environment. When assessing landslide risk at a European scale, Jaedicke et al. (2014) used population density and the density of road and rail networks as metrics to represent exposure. Exposure, and therefore risk, increases when landslides occur in areas of high population or infrastructure density.

Vulnerability is a measure of the potential degree of loss. It is a complex concept as, for example, a well-designed structure could be seen to be less vulnerable than a poorly designed structure, as it is less likely to be damaged in the event of a landslide. In economic terms, however, the well-designed structure may be significantly more valuable than the poorly designed structure, which would make any repairs more costly, meaning economic loss is higher if it is damaged. Vulnerability can also be

[*] k.taalab@ucl.ac.uk
[†] tao.cheng@ucl.ac.uk

considered in terms of individuals, communities and regions, which makes this a difficult metric to measure, especially at the national or international scale.

This study aims to assess the risk that landslides pose to the road and rail network using a case study of Piedmont, Italy (Figure 1A). Italy has been identified as having the greatest extent of infrastructure which is exposed to landslide hazards in the whole of Europe. The Piedmont region is a particularly apt case study as it is within a 'landslide hotspot'- an area of Europe where hazard and risk are greatest (Jaedicke et al., 2014). This study proposes some amendments to the Varnes (1984) risk model as it is applied to landslides. Firstly, that landslide susceptibility should be categorised in terms of the types of landslide that can occur. There are a number of different classes of landslide, based on mass movement characteristics. Each of these will require different mitigation strategies, as well as having different triggers (Cruden, & Varnes, 1996). Secondly, the vulnerability of the road and rail network can be considered in terms of the importance of the road or rail link to the network as a whole. As well as the density of the road and rail network, it is important to assess how critical the exposed parts of the network are. For example, parts of the network are used more frequently can be considered more vulnerable than those less frequently used.

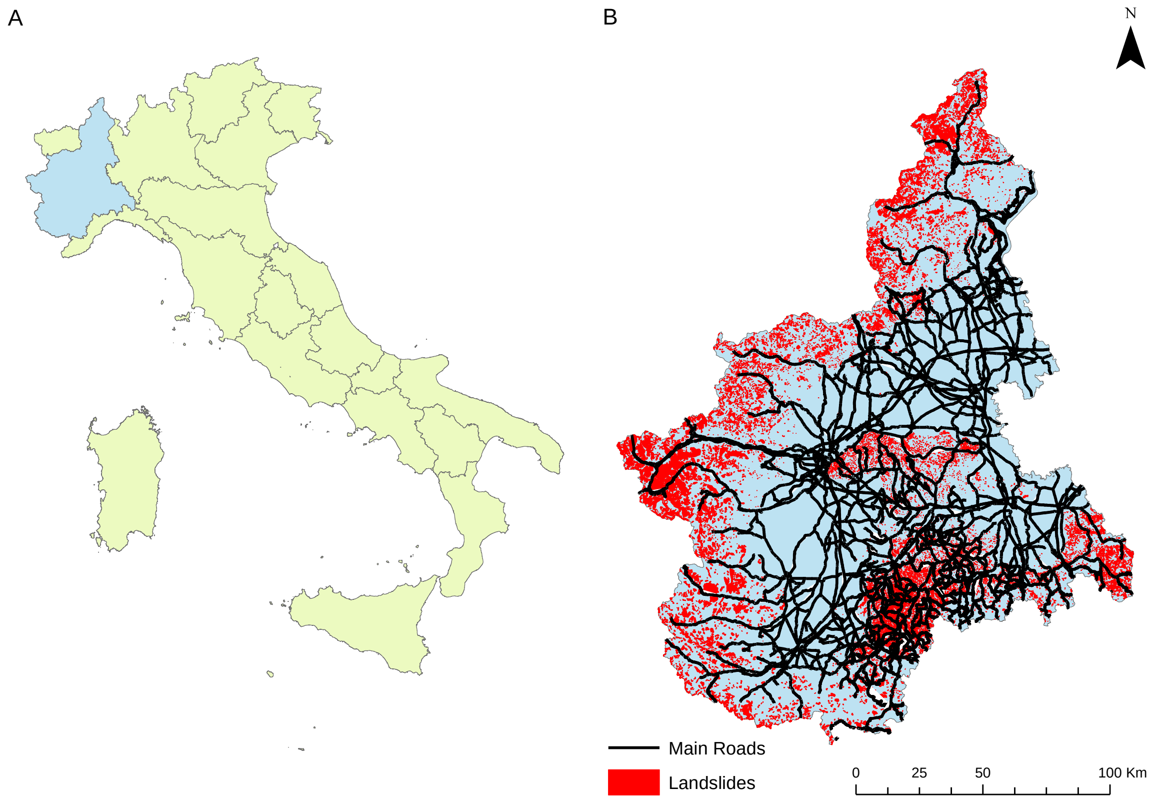

Figure 1 A) Location of Piedmont B) the location of previous landslides and main roads

2. Materials and Methods

2.1 Input data

To develop empirical susceptibility maps, an inventory of previous landslide activity is required. This study uses SIFRAP (the landslide inventory of the Piedmont region) (Lanteri & Colombo, 2013). This dataset contains shapefile records showing the spatial extent of over 30000 previous landslides (Figure 1B). SIFRAP also categorises landslides, a summary of this classification is shown in Table 1.

Table 1 SIFRAP landslide classification

Classification	Description
Crash / Rollover	The mass moves mainly in the air, for free fall, for jumps and rebounds to rolling, shattering into different elements of variable size, and is generally characterized by extremely quick motion
Expansion	An extension movement of cohesive soil or rock, combined with a general subsidence of the mass itself, which fracture and dismantles into several parts, above a soft material, not cohesive
Slow dripping	Movements are generally characterized by low speed and involving soils with high clay content and mostly low water content affecting not very steep slopes
Fast dripping	High speed, affecting mostly loose soils in the presence of significant water content. It is triggered as a result of heavy rainfall and usually involve the loose soil cover on steep slopes
Sliding rotational / translational	Movement along one or more surfaces, where the shear strength is exceeded, or within a zone characterized by relatively thin, intense shear deformation
DGPV	Very complex deformation which occurs through a mostly slow and progressive rock mass, without any appreciable continuous failure surfaces. The process deformation occurs extremely slowly

Other environmental variables used to determine landslide susceptibility are shown in Figure 2. The environmental metrics were selected based on the recommendations of previous empirical studies on landslide susceptibility (Dai & Lee, 2003; Wand & Sassa, 2006; Bui et al., 2013). The 100 m resolution digital elevation model (DEM) was used to derive slope, aspect, curvature and TWI.

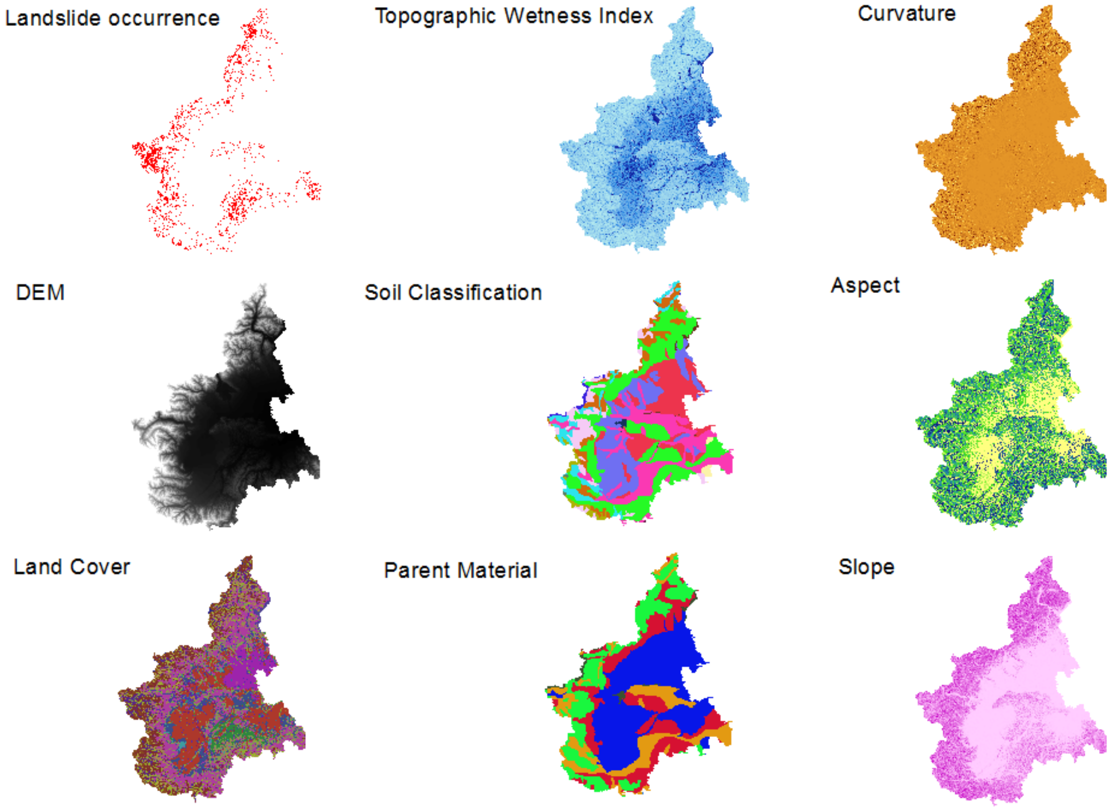

Figure 2 Environmental covariates used for modelling landslide susceptibility

In order to create empirical models of landslide susceptibility, it is necessary to spatially sample all environmental variables at both locations where landslides have occurred previously and at locations where landslides have not occurred previously. This created a dataset of 400000 samples which randomly divided into a training dataset of 300000 samples and a validation dataset of 100000 samples. As well as sampling the values of the environmental variables, samples where landslides were assigned a value of 1 and the landside class was recorded. Samples where no landslide had occurred where given the value 0.

2.2 Support vector machines (SVM)

SVMs are based on statistical learning theory (Vapnik, 1998). Originally developed as a binary classifier, SVMs perform classification (and regression) by constructing N-dimensional hyperplane that optimally separates data into categories (Hearst et al., 1998). For example, if we wish to separate data into two classes, we would like to find a threshold which could discriminate between the two. The simplest example of this would be a straight line in two-dimensional space, or a hyperplane in higher dimensional space. The SVM tries to find the optimal separating hyperplane that gives the largest separation between classes.

Often there is a situation where it is not possible to separate the classes with a straight line or hyperplane. This is there SVMs employ a kernel function (sometimes known as the 'kernel trick'). It is possible to project data which is not linearly separable into higher dimensional space where the data can then be separated by a hyperplane. By using kernel mapping, SVMs can operate in an arbitrary number of dimensions, making it possible to find hyperplane separating solutions for even highly complex datasets (Ballabio & Sterlacchini 2012). Despite the kernel trick, for very complicated datasets, it is usually difficult and not particularly desirable to entirely separate two classes, as this can lead to overfitting. To allow some flexibility, SVMs have a cost parameter C that determines the trade-off between misclassification and enforcing strict (ridged) margins- creating soft margins that allow some misclassification. Increasing the C parameter increases the penalty for misclassifying points, creating a more rigid model.

In high dimensional feature space, support vector regression uses an ε-insensitive loss function to perform linear regression, while reducing model complexity by seeking to minimise $||w||^2$ by solving the optimisation problem shown in Equation 1 (Cherkassky & Ma, 2004)

$$minimise\ \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$
$$subject\ to \begin{cases} y_i - f(x_i,\omega) - b\ \leq \varepsilon + \xi_i^* \\ f(x_i,\omega) + b - y_i\ \leq \varepsilon + \xi_i \\ \xi_i,\xi_i^* \geq 0 \end{cases} \tag{2}$$

Where $||w||$ is the norm of the normal hyperplane, ξ_i^* and ξ_i are slack variable which measure the deviation of the training data beyond the ε-insensitive zone and C is the cost parameter which regulates the relationship between model complexity and error. Overall, SVMs have two parameters which need to be set; the cost parameter C which determines the amount of generalisation and kernel function, which can be linear, polynomial, or Radial Base Function. In this study the kernel function was set as a radial base function due to its robustness (Kavzoglu et al., 2014). The parameter C was determined using the carat package in R and the SVM models were built using the R package e1071 (Meyer et al., 2012).

2.3 Network analysis

Simple network analysis uses the topology of the network to analyse the relative importance of each section of the network. This study proposes using betweeness centrality, which is a measure of how frequently a section of the network is part of the shortest path between points and is used to indicate the influence that a given section of the network will have on flow (in this case traffic flow).
In theory, the higher the centrality, the greater the effect on the network if this section of the network was to be closed, hence the greater the vulnerability in the risk model (Freeman, 1977). This study will used sDNA software for ArcGIS to derive centrality metrics for the road and rail network.

3. Results

Figure 3 shoes the initial result of the SVM landslide susceptibility modelling for Piemont. Given the distribution of pervious landslides, it is to be expected that the areas that show the highest susceptibility are in the west and south of the study area.

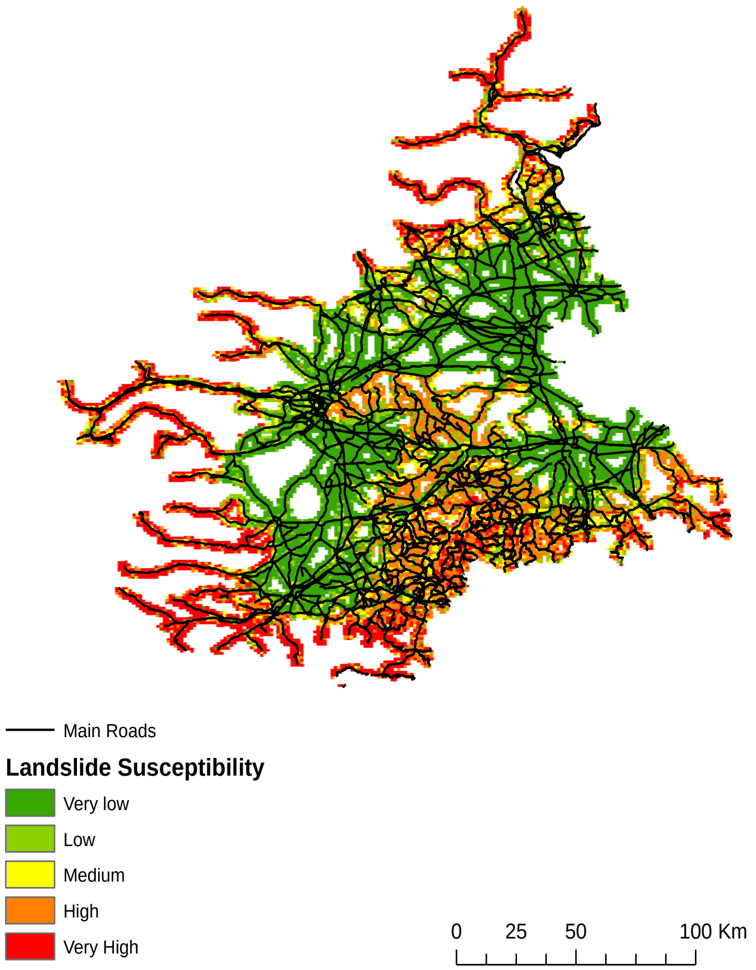

Figure 3 Landslide susceptibility for the road network in Piedmont

4. Future work

In order to quantify the risk landslides, there are a number of tasks which need to be completed

1. Validate the landslide susceptibility map (Figure 1) using test dataset
2. Create a landslide classification map using the SIFRAP data and SVM classification. The data used to train the model will be the environmental data which was sampled in areas where landslides have previously occurred. This will be split into training and validation datasets.
3. Create a normalised betweeness centrality classification for the road and rail network in Piedmont using sDNA software in ArcGIS. This can be used to assess the vulnerability of the network.
4. Using the landslide susceptibility maps as a base, attribute rainfall thresholds to landslide occurrence. Using SVM modelling to develop an empirical relationship between historic rainfall data (both antecedent rainfall and storm events) to identify landslide triggers (e.g. Li et al., 2010; Segoni et al., 2014)

5. References

Ballabio C and Sterlacchini S (2012) Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. *Mathematical geosciences*, 44(1), 47-70.

Bui D T, Pradhan B, Lofman O, Revhaug I and Dick Ø B (2013) Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. *Natural hazards*, 66(2), 707-730.

Cepeda J, Höeg K and Nadim F (2010) Landslide-triggering rainfall thresholds: a conceptual framework. *Quarterly Journal of Engineering Geology and Hydrogeology*, 43(1), 69-84.

Cherkassky V and Ma Y (2004) Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1), 113-126.

Cruden D M and Varnes D J (1996) Landslides: Investigation and Mitigation. Chapter 3-Landslide types and processes. *Transportation research board special report*, (247).

Dai F C and Lee C F (2003) A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic regression. *Earth surface processes and landforms*, 28(5), 527-545.

Freeman L C (1977) A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Hearst M A, Dumais S T, Osman E, Platt J and Scholkopf B (1998) Support vector machines. *Intelligent Systems and their Applications*, IEEE, 13(4), 18-28.

Kavzoglu T, Sahin E K and Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11(3), 425-439

Lanteri L and Colombo A (2013) The Integration Between Satellite Data and Conventional Monitoring System in Order to Update the Arpa Piemonte Landslide Inventory. In *Landslide Science and Practice* (pp. 135-140). Springer Berlin Heidelberg.

Li C, Ma T and Zhu, X (2010). aiNet-and GIS-based regional prediction system for the spatial and temporal probability of rainfall-triggered landslides. *Natural hazards*, 52(1), 57-78.

Meyer D, Dimitriadou E, Hornik K, Leisch F and Weingessel A (2012) *Package: e1071*. R package version 1.6.

Pourghasemi H R, Jirandeh A G, Pradhan B, Xu C and Gokceoglu C (2013) Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *Journal of Earth System Science*, 122(2), 349-369.

sDNA version 1.13.4. Cardiff University. www.cardiff.ac.uk/sdna

Segoni S, Lagomarsino D, Fanti R, Moretti S and Casagli, N (2014) Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. *Landslides*, 1-13.

Varnes D J, The International Association of Engineering Geology Commission on Landslides and Other Mass Movements (1984) *Landslide hazard zonation: a review of principles and practice. Natural Hazards*, vol 3. United nations educational scientific and cultural organization [60 pages]

Vennari C, Gariano S L, Antronico L, Brunetti M T, Iovine G, Peruccacci S, Terranova O and Guzzetti F (2014) Rainfall thresholds for shallow landslide occurrence in Calabria, southern Italy. *Natural Hazards and Earth System Science*, 14(2), 317-330.

Vapnik, V (1998) Statistical Learning Theory. Wiley, New York, NY.

Wang H B and Sassa K (2006) Rainfall induced landslide hazard assessment using artificial neural networks. *Earth Surface Processes and Landforms*, 31(2), 235-247.

6. Acknowledgements

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603960

7. Biography

Khaled Taalab is a research associate in the SpaceTimeLab at University College London. In 2013 he was awarded a PhD in digital soil mapping. He is currently working on a European FP7 project called InfraRISK which is developing models linking natural hazards, critical infrastructure and the environment. His research interests include environmental mapping, data mining and predictive modelling.

Tao Cheng is a Professor in GeoInformatics, and Director of SpaceTimeLab (http://www.ucl.ac.uk/spacetimelab), at University College London. She has broad knowledge and experience in Geographic Information Sciences (GISc), from data acquisition, to information processing, management and analysis, with applications in environmental monitoring, natural resource management, health, transport and crime studies. She has over 140 publications.