Environmental Modelling & Software 76 (2016) 117–127

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Identifying the controls on coastal cliff landslides using machine-learning approaches

Mark E. Dickson*, George L.W. Perry

School of Environment, The University of Auckland, Private Bag 92019, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 27 November 2014
Received in revised form
11 September 2015
Accepted 29 October 2015
Available online 20 November 2015

Keywords:
Landsliding
Machine learning
Erosion
Maxent
Regression trees
Cliffs

ABSTRACT

Transformations are underway in our ability to collect and interrogate remotely sensed data. Here we explore the utility of three machine-learning methods for identifying the controls on coastal cliff landsliding using a dataset from Auckland, New Zealand. Models were built using all available data with a resampling approach used to evaluate uncertainties. All methods identify two dominant landslide predictors (unfailed cliff slope angle and fault proximity). This information could support a range of management approaches, from the development of 'rules-of-thumb' to detailed models that incorporate all predictor information. In our study all statistical approaches correctly predict a high proportion (>85%) of cases. Similar 'success' has been shown in other studies, but important questions should be asked about possible error sources, particularly in regard to absence data. In coastal landslide studies sign decay is a vexing issue, because sites prone to landsliding may also be sites of rapid evidence removal.

1. Introduction

Statistical models are widely used for many different purposes in the earth and environmental sciences. Particularly common are regression methods, which assume an appropriate structural model and then focus on parameterising it. In contrast, machine learning (ML) uses algorithms to learn the relationship between a response and its predictors, and so avoids starting with an assumed structural model (Elith et al., 2008). Many ML techniques have now been developed (Hastie et al., 2009), such as classification and regression trees (CART), maximum entropy models (MAXENT) and boosted regression trees (BRT), which have been used to predict the outcomes of events as diverse as the risk of avian influenza infection (Gilbert et al., 2014), road culvert passability for migratory fishes (Januchowski-Hartley et al., 2014), range shifts in coral-reef habitats under global warming and ocean acidification (Couce et al., 2013), and species distributions (as reviewed in Elith and Leathwick (2009)). To date, however, earth scientists have used these tools much less frequently than in the biological sciences, for instance, although some attention has been placed on the identification of landslide susceptible areas on hillslopes (e.g. Convertino et al., 2013; Felicísimo et al., 2013; Korup and Stolle, 2014). In this study we use three widely employed ML methods – CART, BRT and MAXENT – to evaluate and predict spatial patterns of coastal cliff landsliding. We explore whether these techniques hold promise for coastal management applications, and we investigate whether the difficult conceptual issues surrounding the nature of absence data (i.e. preservation bias or sign decay) that have concerned ecologists building species distribution models also apply in an earth sciences context.

Despite being inherently erosive, cliff-top land remains highly valued for building sites, recreational resources and transportation corridors (Griggs 2005; Young et al., 2014). As a result, cliff erosion poses a hazard in many areas through both small-scale rockfalls and larger landsliding events. This hazard has increased over time due largely to shifts in socio-economic factors increasing the density and economic value of cliff-top developments. In the future this situation is likely to be exacerbated by increases in cliff erosion rates driven by factors such as global sea level rise (Walkden and Dickson, 2008; Ashton et al., 2011).

Management of cliff erosion hazards requires useful model-based forecasts (Walkden and Hall, 2005). Physical process-based models are desirable because they allow a dynamic view of erosion under uncertain future conditions (Dhakal and Sidle, 2004; Vorpahl et al., 2012). However, many challenges exist in process modelling, including the need to underpin models with a detailed understanding of the mechanics of cliff failure, and issues associated with providing model predictions at the temporal and spatial scales required by managers. Encouraging developments are underway, arising both from field techniques such as repeat laser

* Corresponding author.
E-mail address: m.dickson@auckland.ac.nz (M.E. Dickson).

http://dx.doi.org/10.1016/j.envsoft.2015.10.029
1364-8152/

scanning, which reveal dynamics such as progressive-upward failure propagation (e.g. Rosser et al., 2007, 2013), and many examples of detailed numerical analyses on the controls on rock slope failure (e.g. Eberhardt et al., 2004). In this context frequently adopted frameworks include continuum finite-element and finite-difference models, which are used for slopes composed of weak rock masses where failure is controlled by the deformation of the intact material or a restricted number of discrete discontinuities (e.g. faults), and discontinuum techniques that are often used where jointing is the controlling influence on complex rock slope deformation (Stead et al., 2006). However, these studies are usually highly local; typically in the order of a single landslide failure. Managers also require forecasts of failure likelihood over much larger spatial extents and longer durations. To date, on cliffed coasts the use of process-based models for management at these extended space-time domains is limited to cliffed coasts composed of clay, glacial tills, and terrace deposits, where rapid erosion rates provide a historical record of shoreline recession that can be used for model evaluation (e.g. see Dickson et al., 2007). Unfortunately, historical records are often not available for rock coasts composed of more consolidated materials (e.g. sandstones), where erosion may be imperceptibly slow for long periods, interrupted by sudden landsliding failures that can remove several metres of cliff top in a single event. As yet process-models representing the many factors that affect the dynamics and stability of harder-rock cliffs are not available at the spatio-temporal scales (decades, km's) required for coastal management (Dickson et al., 2009).

Lee et al. (2001) discuss the periodicity of landsliding on cliffed coasts where episodic cliff failure events are associated with cliff response to predisposing factors, such as profile steepening by wave action, and triggering factors, such as storms and heavy rainfall. The relationship between these factors is complex: triggering events of the same magnitude may not necessarily lead to landsliding, because preparatory factors may also be required. Such process synergies suggest that successive cliff landsliding events are not independent, because they are influenced by previous events (in other words there are reciprocal feedbacks between pattern and process). Hence, in addition to the scale-limitations on deterministic mechanistic models, traditional statistical models are also not well suited to coastal cliff landsliding.

Limitations in traditional models for coastal cliff erosion are slowly being offset by advances in our ability to collect and interrogate remotely sensed data. For instance, Michoud et al. (2014) describe a boat-based LIDAR survey of a 30 km stretch of cliffed coast in Normandy, France. Such datasets can be analysed by statistical (empirical or data-driven) models, offering alternate, yet complementary, approaches to process-based representations of erosional processes on cliffs. Several modelling methodologies have been explored, including correlative multivariate regression methods (Marques et al., 2013), probabilistic models for generating maximum likelihood distributions of cliff failure (e.g. Hall et al., 2002; Milheiro-Oliveira, 2007), and Bayesian networks to predict spatial variability in the amount of cliff erosion (Hapke and Plant, 2010). Such data-driven approaches are all influenced by the quality and availability of historical data, but developing robust descriptions of long-term change on cliffed coasts is challenging due to the brevity of historical records and monitoring data relative to erosion rates. However, many disciplines face problems arising from the scarcity and patchiness of long-term data records and inventories. The need to make inferences and forecasts under such conditions has resulted in the development of statistical techniques, many grounded in ML approaches, designed to explore messy, non-linear and non-additive data (Hastie et al., 2009; James et al., 2013). Examples of such techniques include CART, MAXENT and BRT, all of which have been widely used in ecological studies (e.g. Elith et al., 2006; De'ath, 2007, Bradley, 2010; Perry et al., 2012).

ML methods can be used both to predict and to make inferences, with one potentially informing the other (James et al., 2013). ML-based techniques have been applied in the earth sciences to identify the relative importance of the potential predictors of landsliding patterns on hillslopes, and thus identify landslide-susceptible areas (Brenning, 2005; Convertino et al., 2013). Felicísimo et al. (2013) compared the performance of four methods (multiple logistic regression, multivariate adaptive regression splines, CART, and MAXENT; the latter three are ML-based) using a landslide database from Spain and concluded that all yielded similarly reliable predictions. However, one issue that has received little attention is the deceptively difficult question of what an 'absence' in a geomorphological dataset really is, and what can be inferred from it (Korup and Stolle, 2014). The issue of absences arises wherever detection is not perfect and, as Lahoz-Monfort et al. (2014) point out, a failure to adequately consider the nature of absences can result in a model predicting detectability rather than presence. No level of statistical sophistication can 'magic away' the issues associated with unreliable parameterisation information (see Lobo, 2008).

We use CART, BRT and MAXENT models to explore spatial patterns in the risk of coastal cliff landslides around Auckland, New Zealand (NZ). This study represents the first application of these techniques to coastal cliff landsliding events. Our main objectives are to: (1) discern the relative importance of the factors that underlie the observed landsliding patterns, and (2) develop statistical models that can be used to predict the spatial pattern of landslide activity. Ultimately our analyses allow us to evaluate the utility of ML-based methods for coastal cliff erosion management, and to contribute to a broader ongoing discussion of presence-absence data in empirical modelling.

2. Field setting and landslide database

Our study is underpinned by coastal landslide data from approximately 40 km of cliffed coastline around Auckland, NZ (Fig. 1). The area encompassed by the database includes cliffs composed of weak sedimentary rocks (interbedded sandstones and mudstones) that have been subject to increasing urban development over several decades, driven in part by population growth in the city that has risen at nearly double the national rate since 1991 (Edbrooke et al., 2003). The cliffs are exposed to limited wave fetch and long-term cliff erosion rates are relatively slow (<0.1 m.yr^{-1}) (de Lange and Moon, 2005). However, sudden episodic cliff failures can remove several metres of cliff top in single events, threatening coastal properties and people (Jongens et al., 2007).

Data collection was funded by Auckland Council and led by a coastal geomorphologist (MD) at the University of Auckland in Feb–July 2010. Data collection involved desk-top mapping using a combination of rectified aerial photographs (2006) and LIDAR-derived contour data (2008) as well as an extensive field mapping programme in 2010. The data collection techniques are summarised in Supplementary Data 1. There is no overt geographic survey bias in the database: mapping was conducted along approximately 40 km of cliffed coast (Fig. 1) within the metropolitan urban limits of the city where the cliffs are composed of sedimentary rock (i.e. omitting hard volcanic rock cliffs). Small sections of coast were omitted where it was not possible to access the cliff toe, and mapping was not conducted on the Manukau Harbour shoreline or on offshore islands. Notwithstanding these restrictions the large range of sites sampled represents a broad environmental coverage with respect to variables analysed.

Initially 64 landslides were located from photographs and contour data, but this represented a considerable under-sampling