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

A R T I C L E I N F O

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

A B S T R A C T

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.

© 2015 Elsevier Ltd. All rights reserved.

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/© 2015 Elsevier Ltd. All rights reserved.

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

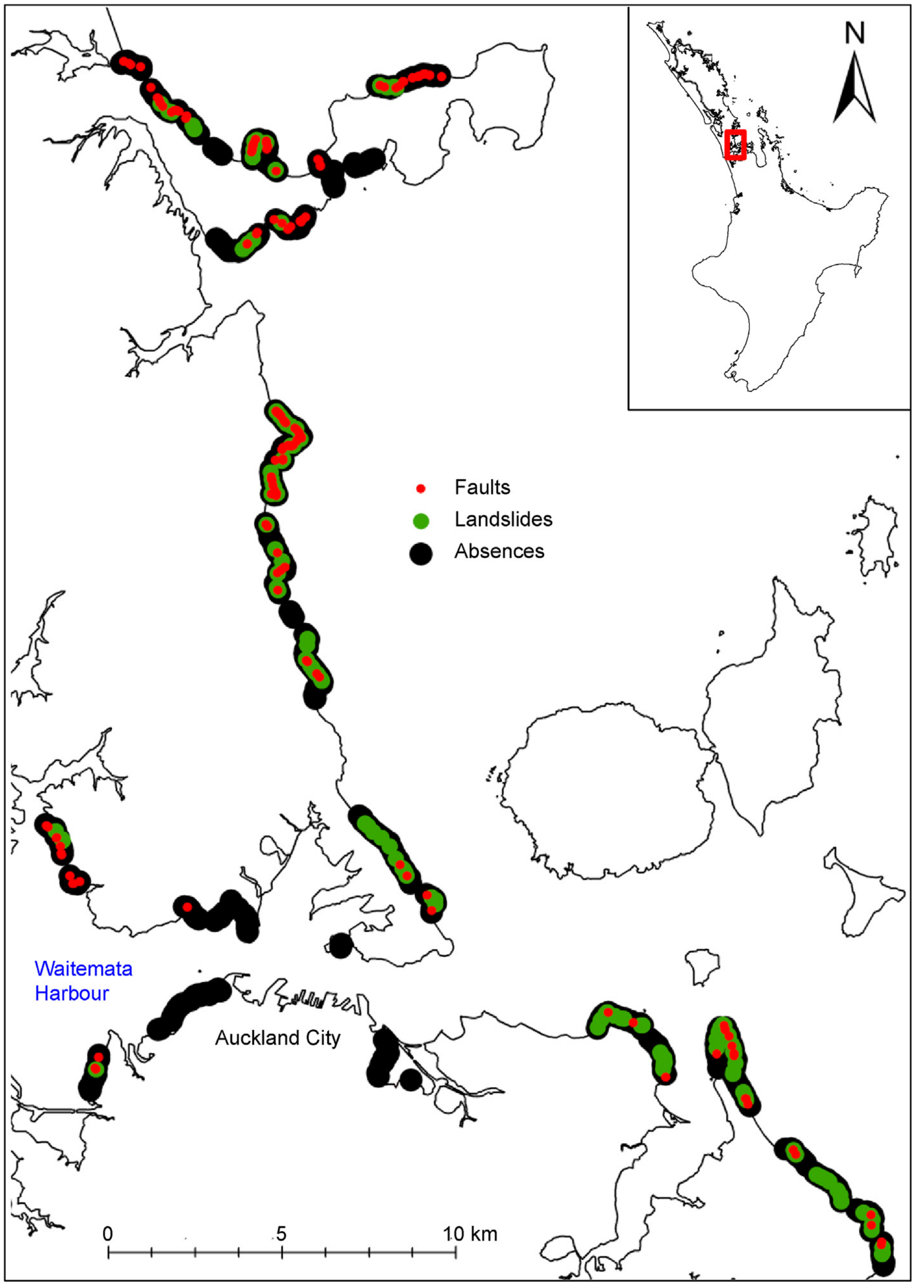

Fig. 1. Map showing sections of cliffed coast around Auckland, NZ in which data collection was undertaken.

due to mapping difficulties associated with vegetation cover and photograph resolution relative to landslide size. A subsequent field mapping programme recognised more than 100 further landslides. These positions were re-inspected using remotely sensed data, and 67 landslides of the >100 landslides recognised in the field were also apparent from photograph or contour data and were added to the database; in total 131 landslides were mapped. There are multiple practical difficulties in compiling such a database, but of particular importance for coastal cliff landslides is a potential invisibility issue that arises because the detection probability is obviously high for sites where landslide evidence is still present, but the evidence for landslides disappears over time (sign decay), and the rate of removal is spatially variable.

In addition to the landslide data points (presence data), information was collected at sites without landslides. The cliff-line was segmented at 100 m increments and mapping was conducted both in the field and from the desk-top at each one of the segments (Fig. 1). In total the database comprises 498 records, 131 presences and 367 absences (prevalence = 0.263). In addition, field inspection indicated that the location of faults in the cliff face represented a likely important control on landsliding. For this reason, during the course of field mapping the location and characteristics of major (i.e. continuous) faults were recorded, producing a database of 162 fault locations.

3. Statistical analyses

3.1. Predictor information

We considered eight predictor variables in our statistical analyses. Table 1 describes the predictors and outlines the method of data acquisition for each; for example, 'unfailed slope' refers to the slope of the cliff face as derived from LIDAR data. Where landslides have occurred the slope is taken immediately adjacent to the landslide, because landslides can impact the slope of the cliff face. The list of predictors is not an exhaustive catalogue of all possible factors controlling cliff failure; rather it is based on extensive qualitative inspections of the field site, reviews of previous studies, and practical limitations of the mapping programme. Six predictor variables reflect attributes of the cliff geomorphology and geology, and two predictors were included to investigate aspects of the process environment (aspect and wave exposure). The variables in Table 1 were used to predict landslide 'presence' (a binary categorical variable). Visual assessment indicated that the predictors were not strongly collinear (highest absolute correlation coefficient among continuous predictors of 0.38; see Supplementary Data 2) so we included all of them in our model building. The sample size (n = 498) and prevalence (0.263) are comfortably inside the adequate levels suggested by Jiménez-Valverde et al. (2009).

3.2. Model implementation

We developed CART, BRT (both presence-absence) and MAXENT (presence-background) models for the full data set and used a resampling approach to assess uncertainty in the model predictions. These three approaches are reviewed by De'ath and Fabricius (2000), Elith et al. (2008, 2011), respectively. The CART and BRT models make the assumption that sites where landslides were not recorded are absences and therefore do not consider the possibility that there was a landslide at a site in the past but it is now obscured or the evidence removed. In contrast to this approach, MAXENT models use only definite presence information and are referred to as presence-background models (see Elith et al., 2011). Note that while CART and the BRT models provide estimates of the probability of an event being present, MAXENT estimates the proportion of landslide event probabilities using a factor that is not identifiable without external data; hence, as Phillips et al. (2006) and Guillera-Arroita et al. (2014) emphasise, MAXENT provides only relative suitability (logistically transformed to a 0–1 scale), which in our case indicates susceptibility of landsliding.

Full data models used all available data (i.e. all 498 sites) to build and evaluate CART, BRT and MAXENT models. While this approach takes advantage of all available data, the data used to build the models must then also be used to evaluate them, which is not ideal. As with many studies we do not have sufficient samples to split the dataset into training and calibration sets. For this reason we also adopted a resampling approach in which a fraction (50%) of the data were selected at random without replacement, a CART/MAXENT/BRT model built from that subset, and that reduced data model then used to predict across the entire dataset. We repeated this process 1000 times, yielding 1000 predictions for each site and 1000 influence estimates for each predictor for each model. This resampling approach is based on that developed by Perry et al. (2012), and allows us to assess the uncertainty attached to predictions at each site and the relative importance of each of the eight predictor variables.

3.3. Model structure and settings

For the CART models we initially developed classification trees including all eight predictors before pruning them using a cost-complexity approach (Hastie et al., 2009). For the BRT models we used a Bernoulli error structure, with learning rate = 0.005, bag fraction = 0.75, and tree complexity = 5. These values produced ensemble BRTs based on a median of 750 (5th and 95th percentiles of 550 and 1100) individual trees, close to the rule of thumb outlined by Elith et al. (2008). For the MAXENT models we used 500 iterations with the convergence threshold set to 1×10^{-5} and the default prevalence to 0.5 (i.e. default settings), and included the few sites where there were missing data for some predictors. If we assume that all absences are true (i.e. we ignore any potential sign decay) we could estimate the true prevalence for the MAXENT model. However, in the spirit of a presence-background model, where prevalence is unknown (Elith et al., 2011), we do not use this information (analyses, not shown here, suggest that using the observed prevalence has little qualitative effect). All statistical analyses were performed using R-3.1.0 (R-Development-Core-Team, 2014) and the dismo (Hijmans et al., 2014) and caret (Kuhn,

Table 1
Summary of predictors used in statistical analyses.

Geology and geomorphology predictors	Description/rationale	Mapping method and data type	Data type	Data summary (min, average, median, max)
Cliff height, metres above mean sea level	The height of cliffs influences shear stress magnitude (Wolters and Müller, 2008). Some studies have used cliff height in hazard zone assessment (e.g. Marques et al., 2013).	LIDAR	Continuous	−0.5 || 21.86 || 20.00 || 60.00
Lithology	Cliffs in the study area are mainly composed of interbedded sandstones and mudstones, but a harder massive unit (Parnell Grit) outcrops at some locations.	Field: distinction between sandstone/mudstone and Parnell Grit.	Binary	
Dip of bedding in degrees	Steeply dipping bedding has been associated with failure (e.g. Trenhaile, 1987).	Field: geological compass.	Continuous	0.0 || 8.0 || 12.8 || 88.0
Direction of bedding dip (seaward or landward)	The direction of bedding influences failure type; seaward sloping bedding is often associated with sliding failures in coastal environments (see Trenhaile, 1987).	Field: distinction between landward or seaward dip.	Binary	
Unfailed slope in degrees	Slope is important in most landsliding studies (see Korup and Stolle, 2014).	LIDAR: at landslide locations the unfailed slope angle was taken immediately adjacent to the landslide.	Continuous	0.00 || 52.30 || 51.88 || 82.23
Proximity to faults in metres	Fault planes represent zones of weakness. Wedge failures are common where faults coincide with joint and bedding planes.	Field: fault locations were mapped; proximity to landslide and non-landslide locations was calculated in ArcMap.	Continuous	0.00 || 92.19 || 333.80 || 4581.00
Process predictors				
Aspect	Aspect might influence cliff erosion rate through differential weathering rates.	Aspect was measured using ArcMap and transformed to values between 0 and 1 (0 = N-facing, 1 = S-facing, 0.5 = E and W, see Cutler et al. 2007)	Continuous between 0 and 1	0.001 || 0.302 || 0.391 || 1.000
Exposure to wave energy; fetch distance in km.	Wave exposure influences wave energy delivery to the cliff toe, which could influence cliff erosion rate and the rate of evidence removal (sign decay).	Fetch provides a simple measure of wave exposure; fetch was calculated using the algorithm described by Burrows et al. (2008).	Continuous	42.70 || 520.30 || 611.30 || 1448.00

2008) packages, and for the MAXENT models, the MAXENT 3.3.3 software (Phillips et al., 2006).

3.4. Model evaluation

For both the 1000 models based on sub-sampling and the models built with the full dataset we estimated the probability value at which converting the prediction for each site from a probability to a presence or absence maximised the kappa statistic (Jiménez-Valverde and Lobo, 2007). We then assessed each model's performance using metrics derived from cross-tabulations of predictions vs. observations (confusion matrices) (Altman and Bland, 1994). We assessed the influence of each predictor on the model predictions by assessing how much the model's performance is influenced by the omission of that predictor (e.g., this forms the basis of the relative contribution (%) metric used to summarise BRT models). For the full model we have a single estimate of relative importance for each parameter but for the resampled models we have 1000 estimates per model and so can estimate the level of uncertainty attached to each. For the BRT and MAXENT models we used univariate partial dependence plots to depict the relationship between each predictor and the predicted probability of landsliding at a given site (Elith et al., 2008).

4. Outcomes

The overall outcome of our CART and BRT (presence-absence) and MAXENT (presence-background) models is summarised in Fig. 2 (see also Table 2 and Supplementary Data 3). All three modelling approaches suggest that the unfailed slope angle of the cliff and proximity to faults are the most important predictors of landslides. This result accords with many qualitative observations of landsliding in the study area, as demonstrated in Fig. 3 which shows two examples where faults in a steep cliff face have led to landsliding. Prior to this study the combined importance of these factors had not been quantified, nor compared against other factors. The overall influence (%) on landsliding respectively attributed to unfailed slope angle and distance to fault is 38.6 ± 5.0 and 37.4 ± 7.5 [CART], 32.9 ± 6.5 and 27.8 ± 6.9 [GBM] and 44.6 ± 8.9 and 41.4 ± 9.0 [MAXENT]; (mean ± 1 SD across the 1000 resampled simulations – see Supplementary Data 3 for full details). In most cases the mean importance values across the 1000 resampled models are close to those of the full model estimates but there is considerable variability in some, which emphasises the utility of the approach we took to model building and uncertainty identification.

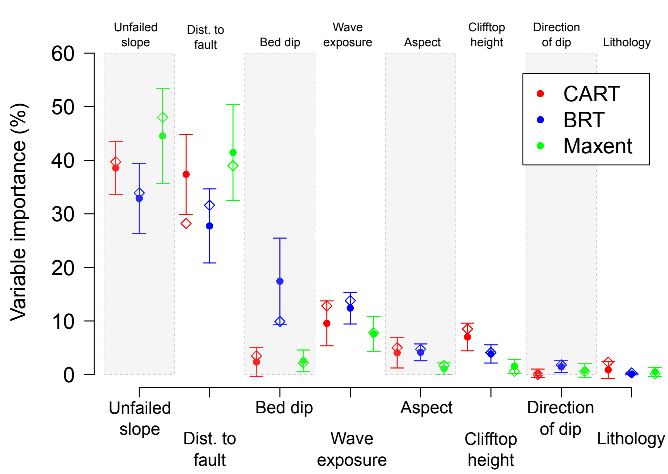

Fig. 2. Overall importance of predictors in the three statistical models used; filled circles and error bars are mean ± 1 SD of the importance values across the 1000 resampled models and the open diamonds the importance values for the full model (see text for details).

The pruned CART model reveals additional detail in the combined effect of unfailed slope and proximity to faults (Fig. 4): in the full model, of the sites that are predicted not to have landslides, about 70% (267/384) are characterised by an absence of faults and (relatively) gentle cliff-face slope angles of less than 56°; by contrast, about 70% (77/114) of the sites predicted to have landslides coincide with locations where the site has a fault and an unfailed slope angle of at least 68°. The response curves for unfailed cliff slope and proximity to faults are shown in Fig. 5. The fault proximity predictor sharply decays with distance, which is entirely as expected, because landslides typically occur precisely at fault locations (Fig. 3), and there is no obvious mechanism to explain a remote influence of faults on landsliding. The response curves for unfailed slope show a near linear increase in influence from 40 to 80° for MAXENT, and a concave upward curve from 40 to 75° for BRT. Slopes shallower than 40° are not associated with landsliding and the flattening of the curves at slopes about 75–80° results from a lack of data, because such steep slopes are likely beyond the critical failure slope angle (see Wolters and Müller, 2008).

The relative importance of other predictors varies between the three models used (Fig. 2). The pruned CART model and the MAXENT model each consider exposure to wave energy as the only other important predictor variable (9.6 ± 4.2 and 7.6 ± 3.3%, respectively). In contrast, BRT suggests that the third and fourth most important predictors respectively are the bedding dip of strata within the cliff, and wave exposure (17.4 ± 8.0 and 12.4 ± 3.0%, respectively). It is not surprising that bedding dip is an important predictor, because the rocks in the study area are intensively folded in places, and steep bedding is sometimes associated with landslides (Fig. 6). The shape of the response curve for bedding dip (Fig. 5) is intuitive, as the slope of the bed is increasingly influential up to about 18° (i.e. landsliding is increasingly likely with more contorted bedding), beyond which there are relatively little data. The dominant direction of dip, either toward the sea or cliff, was unimportant in our models (Fig. 2).

All three models recognise wave exposure as an important predictor of landsliding. Exposure to wave energy is intuitively associated with landsliding, because greater wave energy is likely to drive faster rates of cliff undercutting and therefore more frequent landsliding. However, the response curves for wave exposure (Fig. 5) both show peak influence for a mid-range wave exposure of about 400 km fetch-length (around the 33rd percentile of fetch lengths). It is unsurprising that influence increases from highly sheltered sites (exposure close to zero), but the *prima facie* expectation would be a monotonic increase toward the most open-ocean (exposed) sites. The apparent decline in influence of wave exposure for greater fetches may result from greater capacity of waves to remove the evidence of landsliding. If such sign-decay absences exist in our dataset, then these might be more prevalent at more exposed sites. Nevertheless, wave exposure shows qualitatively the same responses, and has similar predictive importance, in both presence-absence and presence-background models.

The remaining predictors had a mean combined influence of only 5.2% for MAXENT, 14.5% for the pruned CART model and 10.9% for BRT, of which the majority was contributed by aspect and cliff height, respectively (Fig. 2). There is no obvious spatial trend in response curves for aspect, whereas predicted probabilities increase with increasing cliff height, perhaps because larger volume landslides occur on higher cliffs and so are more readily detectable in the landscape. Lithology does not have an important influence because the vast majority of the cliffed coast is formed in one

Table 2
Model performance statistics for the three algorithms (see Altman and Bland, 1994 and Kuhn, 2008 for details); PCC = percent correctly classified (overall accuracy of model), 'Optimal p' is the probability threshold that maximises the kappa statistic when the predictions are binarised into present vs. absent and sensitivity and specificity are the true and negative prediction rates, respectively. For the resampled models the values shown are 5th percentile-**median**-95th percentile.

Model	Set	Prevalence	Optimal p	PPC	Sensitivity	Specificity
CART	Full	0.263	0.733	0.890	0.725	0.948
	Samples	0.229–**0.261**–0.297	0.444–**0.762**–0.914	0.831–**0.855**–0.873	0.435–**0.588**–0.756	0.888–**0.948**–0.975
BRT	Full	0.263	0.502	0.953	0.855	0.989
	Samples	0.229–**0.265**–0.297	0.275–**0.399**–0.533	0.890–**0.906**–0.924	0.718–**0.794**–0.863	0.916–**0.948**–0.975
MAXENT	Full	0.263	0.425	0.931	0.733	0.959
	Samples	0.229–**0.261**–0.293	0.387–**0.434**–0.485	0.855–**0.873**–0.886	0.595–**0.679**–0.756	0.901–**0.944**–0.978

Fig. 3. Photographs showing steep cliffs where landsliding failure is associated with faults. Note that talus is apparent in photograph (a) but not in (b).

lithology class (interbedded sandstone/siltstone), and the majority of landslides also occur in that class (Supplementary Data 2).

The overall performance of the three models for predicting landslide occurrence is summarised in Fig. 7 and Table 2. The predictive accuracy of all of the models is high (in excess of 85% for all), but in the typical range reported by Korup and Stolle (2014) for ML predictions of landsliding. On the basis of confusion matrix statistics the BRT models slightly outperform the other two approaches (Table 2). Across all three models, specificity (the rate of true absences) is consistently higher than sensitivity (the rate of true presences), and for two of the three, the negative prediction value is slightly higher than the positive prediction value. These trends suggest that the models' ability to predict absences is slightly better than it is for presences. The performance of the resampled models is slightly worse than that for the full model but the metrics for the full model consistently lie inside the 95th percentile of those for the 1000 resampled predictions (Table 2).

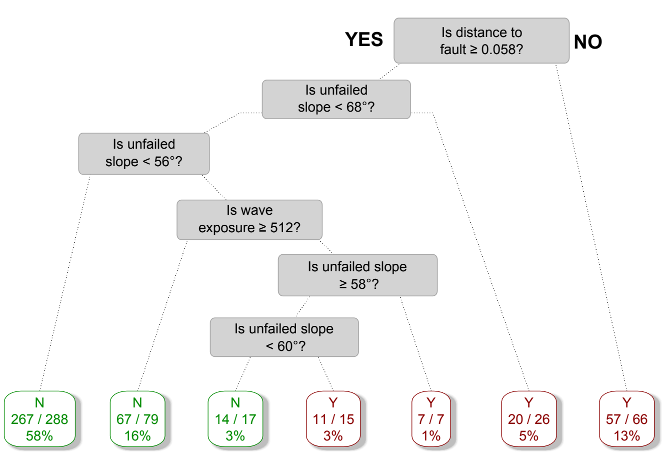

Fig. 4. Pruned CART model, highlighting the importance of fault locations and unfailed slope angle as predictors of erosion activity. Values under each terminal node are the number of correct cases, the total number of cases and the percentage of the data-set falling into each terminal node (so for the far left node, 267 of the 288 cases were correctly classified and this accounts for 58% of the data [288/498]).

5. Discussion

The discussion below is broadly framed around prospects for ML-based statistical techniques in an era of burgeoning earth-science data (Section 5.3). Focus is provided in Section 5.1, which considers the potential utility of the techniques on a relatively small coastal landslide database, and Section 5.2 which discusses some of the basic limitations that landscape data impose on data-driven models.

5.1. Management of coastal cliff landsliding

Faced with sea-level rise and multifarious human impacts on coastal areas, shoreline management plans have begun to adopt longer-term planning horizons (several decades). Such plans need

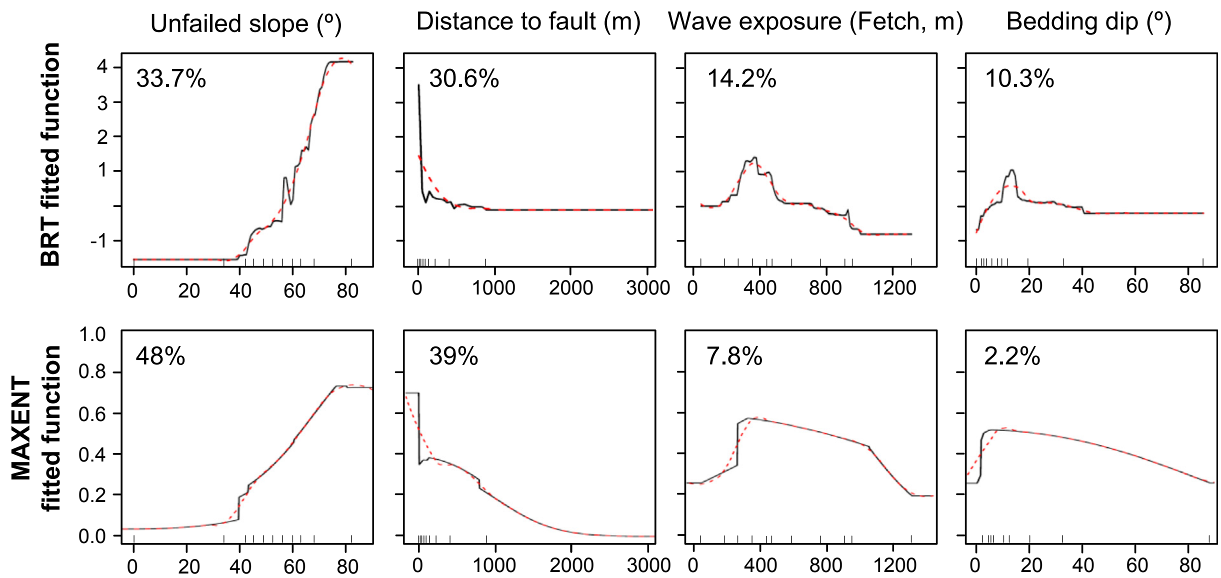

Fig. 5. Univariate partial dependency curves (the effect of a given variable with all others factored out) showing fitted functions (expressed as marginal effects from BRT and MAXENT models for the four most important predictors with the relative importance (full model) of each of the four parameters indicated; black lines are raw values and dashed lines smoothed curves. The dashes on the x-axis are a rug plot and show the distribution of the values in the predictor dataset.

Fig. 6. Photograph showing steep bedding and a landslide.

to be informed by useful model predictions. Progress has been made on cliffed coasts composed of glacial tills and clays using process-based models that are calibrated against historical erosion data (e.g. Dickson et al., 2007). By contrast, models of harder rock shores have focussed on improving scientific understanding over geological timescales (e.g. Trenhaile, 2000). Ironically, one of the management problems associated with slow (i.e. cm.yr^{-1}) erosion rates on rocky shores is that such low rates of change lessen perceptions of risk and encourage landowners to build close to the cliff top, yet episodic landsliding events occur and can remove metres of cliff top in single events. Slow rates of erosion also pose a challenge for modellers because few data are available with which to help implement, parameterise and evaluate process-based models. At present there are no numerical models for slowly eroding coasts that are capable of prediction across the space-time scales of most use to managers.

A principal outcome of our study is that the ML techniques we used were successful in revealing fundamental relationships that influence coastal cliff landsliding from a relatively small database ($n = 498$, c.f. $n > 27{,}500$ landslide observations in the Arno basin, Italy, Convertino et al., 2013) compiled during a single programme (snapshot) of data collection. All three of the techniques that we used had high predictive accuracy (greater than 85%) and, more importantly, consistently identified the unfailed slope of the cliff face and proximity to faults as the most important variables for explaining observed patterns of landsliding. A logical next step in the application of ML approaches to geomorphological hazards is to use ensemble approaches as has been performed for species distribution models (Araújo and New, 2007). While other predictors, beyond those we considered, might be important, the results of our analyses are intuitive and draw attention to the primary importance of geological and geomorphological predictors across our field site. However, as we consider below and as Korup and Stolle (2014) caution, such high success rates require careful interpretation: in the context of coastal cliff landsliding, absences arising from sign decay may be particularly important.

Having characterised the conditions under which landsliding failure occurs, it is possible to build models in which the predictive outcome are maps that show the spatial distribution of landsliding. The resampling method we adopt means we can accompany these

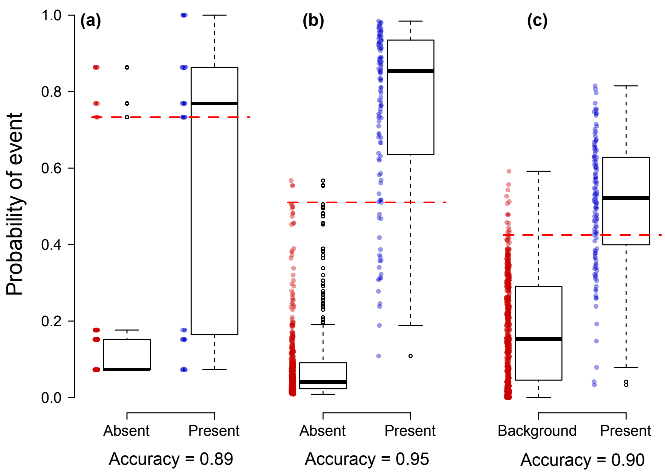

Fig. 7. Boxplots showing probabilities of cliff landsliding for (a) CART, (b) BRT and (c) MAXENT models for presences and absences (a and b) and presence and background (c). Points show raw probability values for individual cases (jittered for visual clarity) and the horizontal dashed lines the probability threshold that maximises the kappa statistic when probability values are binarised into presence vs. absence. Note that the recursive partitioning that the CART approach uses results in a set of discrete predictions.

maps with their associated uncertainties. Such predictions are not definite statements about what will happen in the future; rather, via reverse inference, they depict the relative likelihood of landsliding on the basis of historical patterns of such activity.

In Fig. 8 we provide an example of an arbitrary section of shoreline in which we have used a BRT model with resampling (see Section 3.2) to predict the relative spatial distribution of landslide likelihood, and the associated uncertainty (represented in this case by the range of data). Data such as these might be used by managers to calculate setback lines. Setbacks have a long history in coastal management (e.g. Shows, 1978) and are intended to prevent or restrict development. However, their effectiveness has been lessened by a combination of political processes (see Abbott, 2013) and technical issues around appropriately defining setback distances (e.g. see Moon and Healy, 1994 and Jongens et al., 2007 in the context of cliffed coasts). In Fig. 8 a spatially variable setback is schematically represented with a red hatched line that runs closer to the cliff top in areas where landsliding likelihood is lower. Undoubtedly there would be many political issues to consider in this

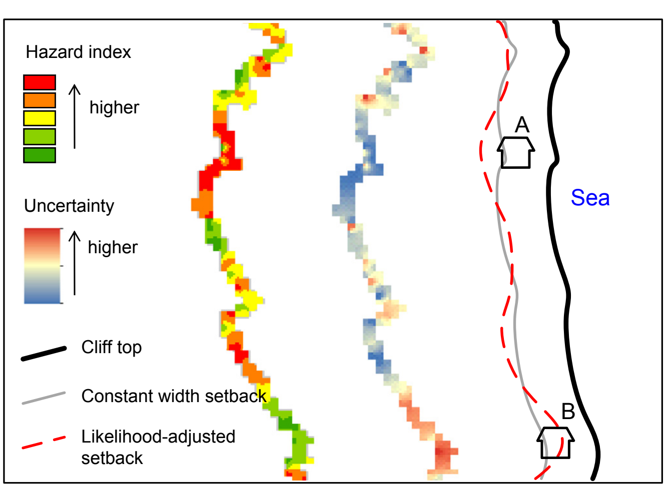

Fig. 8. An example of spatial landsliding likelihood mapping to illustrate possible hazard mapping utility.

hypothetical example. For instance, two houses with equivalent risk in a constant-width setback scenario are alternately placed inside and outside of a likelihood-adjusted setback, but the owner of House B would point toward high uncertainty associated with model predictions near their property. Faced with such uncertainty, managers might choose to adopt simpler data-driven heuristics for cliff-top planning. For instance, the pruned CART model (Fig. 4) suggests threshold conditions that could represent useful rules-of-thumb for managers. For the coastline we have studied it is reasonable to conclude that sites with high (relative) likelihoods of landsliding occur on slopes of at least 68° and where faults intersect the cliff face. A first-order, but useful management exercise could involve using geological maps and topographic maps to locate sites of potential concern. These sites could form locations for more detailed geotechnical analyses, and perhaps stabilisation projects.

5.2. *Presence/absence of coastal landslides*

Ecologists have considered absences in the context of species distribution modelling and Lobo (2008) found that the use of appropriate absence data can significantly improve the performance of such models. Lobo et al. (2010) discriminate between three types of absence: (i) contingent absences, which are related to the inability of a species to access suitable habitat, (ii) environmental absences, where biophysical conditions are not conducive to the event occurring, and (iii) methodological absences, such as survey bias. These three types of absence can also be seen in the earth sciences. For example, human-engineered hillslopes where landsliding is prevented represent a contingent absence; topography that is too shallow to permit landsliding represent an environmental absence; and survey bias (e.g. inaccessibility of some sites) is as problematic for earth scientists as for ecologists. In addition to these, a fourth type of absence, sign decay, occurs where static evidence has been removed or is otherwise no longer present. In ecology, sign decay occurs where the evidence for a species is indirect and impermanent (e.g. footprints or scat), taking the form of a physical mark on the landscape (as opposed to direct sightings). In the earth sciences, sign decay, and attendant ‘invisibility’, may be a prevalent issue, because geomorphological processes vary in time and space and constantly modify and overwrite the landscape. One example is provided by archaeological sites on hillslopes that are subject to varying rates of soil erosion (Wainwright, 1994).

Coastal cliff landslides are also subject to sign decay, and in a particularly vexing manner. Absences occur on cliffed coasts at locations where the evidence of the previous landslide has been removed. As such, they cannot be interpreted as locations where a landslide has never occurred, because over sufficiently long time scales all sites along a cliff face are subject to erosion. Absences might arise from environmental conditions that are not amenable to landsliding, or they may result from a high rate of sign decay (evidence removal). Conversely, presences might arise either from environmental conditions that are amenable to landsliding or where there are very low rates of sign decay. This situation is made more complex when cliff wave exposure is considered, because high exposure might be associated both with high landslide frequency (through cliff undercutting), and also high rates of sign decay (through rapid removal of landslide evidence). By contrast, low wave exposure might be associated with lower landslide frequency, but also with lower rates of sign decay, which acts to increase the chance of presence observations. Similar issues have been noted by ecologists (e.g. see Cristescu et al., 2012), and it is clear that pattern-matching approaches, such as that developed here, need to be clearly placed in the context of the temporal depth of the data used to inform them. However, overall we found little difference in the predictive performance or predictor identification of presence-absence and presence-background approaches. It is possible that because variability in wave exposure simultaneously associates higher/lower event frequency with higher/lower sign decay rates, error has been (fortuitously) minimised. Our model results are qualitatively sensible and represent a substantial contribution to the coastal cliff management toolbox, but appropriately handling absence data and sign decay is an active area of research (see Rhodes et al., 2011; Barbet-Massin et al., 2012; Hertzog et al., 2014) and one that is likely to have important implications for many earth science applications.

5.3. Machine-learning analyses of 'big' earth-science data

The earth sciences are experiencing a 'data deluge' (Mattmann, 2013), with aerial and satellite images now providing near continuous observations of the earth's surface. In this setting data-mining algorithms are invaluable, because they help scientists learn or discover unknown properties and patterns contained within massive data sets (Sellars et al., 2013). At present, the data deluge is mainly 2D, derived from aerial photographs and satellites, but the speed of acquisition of 3D data is being transformed through techniques such as airborne LIDAR, ground- and boat-based laser scanning (e.g. Rosser et al., 2007, 2013; Michoud et al., 2014), and structure-from-motion (e.g. Westoby et al., 2012; James and Robson, 2012; Fonstad et al., 2013; Micheletti et al., 2014). For instance, it is now possible using drone-mounted cameras and structure-from-motion software to compile high resolution (cm-scale) 3D digital terrain models over large areas (tens of km's) quickly and relatively cheaply. In this context, what are the prospects for the type of ML-based techniques that we describe?

The coastal landsliding database from Auckland analysed in this paper was collected over several months and involved extensive field survey. That data collection task could now be largely automated using a drone. Coverage would improve, because field surveyors could physically not access some coastal areas due to tidal inundation. With drone-based data collection, field mapping of landslide predictors (e.g. bed dip, fault location) would be replaced by digital mapping in a Geographical Information System. This would be significantly faster than field mapping, therefore enabling mapping at higher resolution, and generation of a larger database. Digital mapping would no doubt introduce error sources, but inaccuracies also occur in field sampling. For instance, multiple people were involved in field mapping, each making their own judgements when recording data, albeit constrained by paper templates.

We believe that the statistical methods described in this paper are likely in the future to find broad application in the earth sciences, both in the analysis of existing datasets and also in new mass data acquisition projects. The learning-based methods that we use here have long been recognised as potentially more scalable than 'traditional' regression approaches (Gahegan, 2000). There is increasing emphasis being placed on their computational scalability to take advantage of, for example, distributed and parallel computing architectures and they can now be applied to massive datasets (orders of magnitude larger than that which we consider). However, important limitations exist. In particular, there are issues of spatial transferability: to what extent can a model trained on data from one location be used to predict outcomes at another? Clearly there is heterogeneity in the landscape, such as variability in lithology and rock structure, that limit the breadth of application from site to site. This is, of course, an important limit on all pattern-matching modelling approaches. However, data-driven models are not just 'predict-and-prescribe' (Schindler and Hilborn, 2015, p. 953); they can also be used inferentially. For example, statistical models can be used to inform process-based approaches by providing a filter on the key controls at play in the scales at which they operate, and this type of framework has successfully been used in the development of land-use change models (e.g. Millington et al., 2008). As with any data-driven approach the quality of the data are key to the inferences that can be made from them. We have highlighted the need for a careful interpretation of success rates (see also Korup and Stolle, 2014) and the nature of geomorphic absences. These are active areas of research and of relevance across a number of disciplines.

In the context of coastal cliff landsliding, we view ML-based approaches as a complementary addition to the tool-box of methods available to manage coastal cliff retreat. At the scale of individual landsliding events, the best understanding of failure mechanics and drivers will be obtained from direct measurements of ground movement (e.g. Rosser et al., 2007), and modelling studies including discontinuum techniques, continuum finite-element and finite-difference models (see Stead et al., 2006). At larger temporal and spatial scales, historical datasets can be modelled with multivariate regression methods (Marques et al., 2013), probabilistic methods (e.g. Hall et al., 2002), and Bayesian networks (Hapke and Plant, 2010). However, ML-based techniques, augmented with resampling approaches, offer several further benefits for coastal cliff landslide mapping. We found that using a relatively small, not perfectly complete dataset containing varied data types, it is possible to obtain: (i) reliable estimates of the variables that most strongly influence landsliding, without *a priori* knowledge, and (ii) high overall prediction success rates, together with estimates of uncertainty, that can inform spatially variable landslide hazard maps.

6. Conclusions

This study provides the first application of machine-learning based statistical techniques to coastal cliff landsliding hazards. Useful results were obtained from analyses of a relatively small dataset comprising 498 landslide presence and absence observations from the cliffed coast of Auckland, NZ. Three methods were used: classification and regression trees, boosted regression trees, and maximum entropy models. All methods identified the same two predictors (unfailed cliff slope angle and proximity to faults) as dominant in explaining the observed landsliding pattern. This information can be used to derive useful heuristic tools for managers. For instance, high relative likelihood of landsliding generally occurs at sites where faults intersect the cliff face and where slope angle is at least 68°. When all predictor information is used, each of the three statistical approaches predicts a high proportion (>85%) of cases correctly. Hence, it is possible to generate landsliding likelihood hazard maps, together with uncertainty estimates obtained through a resampling process. However, important questions remain to be answered regarding possible error sources. Of these, sign decay (preservation bias) is particularly notable at the coast, because higher wave exposure may remove landslide evidence faster, while also promoting higher landsliding frequency.

Acknowledgements

Funding for this study was provided by Auckland Council. We gratefully acknowledge Marc-Andre Brideau for assistance classifying landslides, and Catherine Frericks, Farah Lodhawalla, Rachael Pentney, and Vaughan Turner for assistance in data collection. We gratefully acknowledge the constructive comments of four anonymous reviewers.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2015.10.029.

References

Abbott, T., 2013. Shifting shorelines and political winds–The complexities of implementing the simple idea of shoreline setbacks for oceanfront developments in Maui, Hawaii. Ocean Coast. Manag. 73, 13–21.

Altman, D.G., Bland, J.M., 1994. Statistics notes: diagnostic tests 2: predictive values. Br. Med. J. 309 (6947), 102.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22 (1), 42–47.

Ashton, A.D., Walkden, M.J., Dickson, M.E., 2011. Equilibrium responses of cliffed coasts to changes in the rate of sea level rise. Mar. Geol. 284 (1), 217–229.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3 (2), 327–338.

Bradley, B.A., 2010. Assessing ecosystem threats from global and regional change: hierarchical modeling of risk to sagebrush ecosystems from climate change, land use and invasive species in Nevada, USA. Ecography 33 (1), 198–208.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5 (6), 853–862.

Burrows, M.T., Harvey, R., Robb, L., 2008. Wave exposure indices from digital coastlines and the prediction of rocky shore community structure. Mar. Ecol. Prog. Ser. 353, 1.

Convertino, M., Troccoli, A., Catani, F., 2013. Detecting fingerprints of landslide drivers: a MaxEnt model. J. Geophys. Res. Earth Surf. 118 (3), 1367–1386.

Couce, E., Ridgwell, A., Hendy, E.J., 2013. Future habitat suitability for coral reef ecosystems under global warming and ocean acidification. Glob. change Biol. 19 (12), 3592–3606.

Cristescu, R.H., Goethals, K., Banks, P.B., Carrick, F.N., Frère, C., 2012. Experimental evaluation of koala scat persistence and detectability with implications for pellet-based fauna census. Int. J. Zool. 2012.

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.

de Lange, W.P., Moon, V.G., 2005. Estimating long-term cliff recession rates from shore platform widths. Eng. Geol. 80 (3), 292–301.

De'ath, G., 2007. Boosted trees for ecological modeling and prediction. Ecology 88 (1), 243–251.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81 (11), 3178–3192.

Dhakal, A.S., Sidle, R.C., 2004. Distributed simulations of landslides for different rainfall conditions. Hydrol. Process. 18 (4), 757–776.

Dickson, M.E., Walkden, M.J., Hall, J.W., 2007. Systemic impacts of climate change on an eroding coastal region over the twenty-first century. Clim. change 84 (2), 141–166.

Dickson, M.E., Woodroffe, C.D., Cowell, P.J., 2009. Geomorphological models and their role in coastal management. In: Green, D. (Ed.), Coastal Zone Management. Thomas Telford, London, pp. 157–184.

Eberhardt, E., Stead, D., Coggan, J., 2004. Numerical analysis of initiation and progressive failure in natural rock slopes—the 1991 Randa rockslide. Int. J. Rock Mech. Min. Sci. 41 (1), 69–87.

Edbrooke, S.W., Mazengarb, C., Stephenson, W., 2003. Geology and geological hazards of the Auckland urban area, New Zealand. Quat. Int. 103 (1), 3–21.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC. M., Peterson, A.Townsend, Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29 (2), 129–151.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (1), 677.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Animal Ecol. 77 (4), 802–813.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57.

Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10 (2), 175–189.

Fonstad, M.A., Dietrich, J.T., Courville, B.C., Jensen, J.L., Carbonneau, P.E., 2013. Topographic structure from motion: a new development in photogrammetric measurement. Earth Surf. Process. Landforms 38 (4), 421–430.

Gahegan, M., 2000. On the application of inductive machine learning tools to geographical analysis. Geogr. Anal. 32, 113–139.

Gilbert, M., Golding, N., Zhou, H., Wint, G.W., Robinson, T.P., Tatem, A.J., Lai, S., Zhou, S., Jiang, H., Guo, D., 2014. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. Nat. Commun. 5 http://dx.doi.org/10.1038/ncomms5116.

Griggs, G.B., 2005. In: California's retreating coastline: where do we go from here, California and the World Ocean. Conference, pp. 121–125.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2014. Maxent is not a presence–absence method: a comment on Thibaud et al. Methods Ecol. Evol. http://dx.doi.org/10.1111/2041-210X.12252.

Hall, J.W., Meadowcroft, I.C., Lee, E.M., van Gelder, P.H., 2002. Stochastic simulation of episodic soft coastal cliff recession. Coast. Eng. 46 (3), 159–174.

Hapke, C., Plant, N., 2010. Predicting coastal cliff erosion using a Bayesian probabilistic model. Mar. Geol. 278 (1), 140–149.

Hastie, T., Friedman, J., Tibshirani, R., 2009. The Elements of Statistical Learning - Data Mining, Inference, and Prediction. Springer, New York.

Hertzog, L.R., Besnard, A., Jay-Robert, P., 2014. Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. Divers. Distrib. 20, 1403–1413.

Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2014. Dismo: Species Distribution Modeling. R Package Version 1.0-5. http://CRAN.R-project.org/package=dismo.

James, M., Robson, S., 2012. Straightforward reconstruction of 3D surfaces and topography with a camera: accuracy and geoscience application. J. Geophys. Res. Earth Surf. (F3), 117 (2003–2012).

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer, New York.

Januchowski-Hartley, S.R., Diebel, M., Doran, P.J., McIntyre, P.B., 2014. Predicting road culvert passability for migratory fishes. Divers. Distrib. 20, 1414–1424.

Jiménez-Valverde, A., Lobo, J., Hortal, J., 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Commun. Ecol. 10 (2), 196–205.

Jiménez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta Oecol. 31 (3), 361–369.

Jongens, R., Gibb, J., Alloway, B., 2007. A new hazard zonation methodology applied to residentially developed sea-cliffs with very low erosion rates, East Coast Bays, Auckland, New Zealand. Nat. hazards 40 (1), 223–244.

Korup, O., Stolle, A., 2014. Landslide prediction from machine learning. Geol. Today 30 (1), 26–33.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5), 1–26.

Lahoz-Monfort, J.J., Guillera-Arroita, G., Wintle, B.A., 2014. Imperfect detection impacts the performance of species distribution models. Glob. Ecol. Biogeogr. 23 (4), 504–515.

Lee, E., Hall, J., Meadowcroft, I., 2001. Coastal cliff recession: the use of probabilistic prediction methods. Geomorphology 40 (3), 253–269.

Lobo, J.M., 2008. More complex distribution models or more representative data? Biodivers. Inf. 5, 14–19.

Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33 (1), 103–114.

Marques, F., Matildes, R., Redweik, P., 2013. Sea cliff instability susceptibility at regional scale: a statistically based assessment in southern Algarve, Portugal. Nat. Hazards Earth Syst. Sci. Discuss. 1 (3), 1965–2003.

Mattmann, C.A., 2013. Computing: a vision for data science. Nature 493 (7433), 473–475.

Micheletti, N., Chandler, J.H., Lane, S.N., 2014. Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. Earth Surf. Process. Landforms. http://dx.doi.org/10.1002/esp.3648.

Michoud, C., Carrea, D., Costa, S., Derron, M.-H., Jaboyedoff, M., Delacourt, C., Maquaire, O., Letortu, P., Davidson, R., 2014. Landslide detection and monitoring capability of boat-based mobile laser scanning along Dieppe coastal cliffs, Normandy. Landslides 1–16.

Milheiro-Oliveira, P., 2007. Bayesian statistical methods for modeling and prediction of major landslides in coastal cliffs. Coast. Eng. J. 49 (01), 45–61.

Millington, J., Romero-Calcerrada, R., Wainwright, J., Perry, G., 2008. An agent-based model of Mediterranean agricultural land-use/cover change for examining wildfire risk. J. Artif. Soc. Soc. Simul. 11 (4), 4.

Moon, V.G., Healy, T., 1994. Mechanisms of coastal cliff retreat and hazard zone delineation in soft flysch deposits. J. Coast. Res. 663–680.

Perry, G.L.W., Wilmshurst, J.M., McGlone, M.S., Napier, A., 2012. Reconstructing spatial vulnerability to forest loss by fire in pre-historic New Zealand. Glob. Ecol. Biogeogr. 21 (10), 1029–1041.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3), 231–259.

R-Development-Core-Team, 2014. R: a Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria).

Rhodes, J.R., Lunney, D., Moon, C., Matthews, A., McAlpine, C.A., 2011. The consequences of using indirect signs that decay to determine species' occupancy. Ecography 34 (1), 141–150.

Rosser, N., Lim, M., Petley, D., Dunning, S., Allison, R., 2007. Patterns of precursory rockfall prior to slope failure. J. Geophys. Res. Earth Surf. 112 (F4), 2003–2012.

Rosser, N.J., Brain, M.J., Petley, D.N., Lim, M., Norman, E.C., 2013. Coastline retreat via progressive failure of rocky coastal cliffs. Geology 41 (8), 939–942.

Schindler, D.E., Hilborn, R., 2015. Prediction, precaution, and policy under global change. Science 347, 953–954.

Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K.-l., Sorooshian, S., 2013. Computational Earth science: big data transformed into insight. Eos Trans. Am. Geophys. Union 94 (32), 277–278.

Shows, E.W., 1978. Florida's coastal setback line—an effort to regulate beachfront

development. Coast. Manag. 4 (1–2), 151–164.
Stead, D., Eberhardt, E., Coggan, J., 2006. Developments in the characterization of complex rock slope deformation and failure using numerical modelling techniques. Eng. Geol. 83 (1), 217–235.
Trenhaile, A., 2000. Modeling the development of wave-cut shore platforms. Mar. Geol. 166 (1), 163–178.
Trenhaile, A.S., 1987. The geomorphology of rock coasts. Oxford University Press, USA.
Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B., 2012. How can statistical models help to determine driving factors of landslides? Ecol. Model. 239, 27–39.
Wainwright, J., 1994. Erosion of archaeological sites: results and implications of a site simulation model. Geoarchaeology 9 (3), 173–201.
Walkden, M., Dickson, M., 2008. Equilibrium erosion of soft rock shores with a shallow or absent beach under increased sea level rise. Mar. Geol. 251 (1), 75–84.
Walkden, M., Hall, J., 2005. A predictive mesoscale model of the erosion and profile development of soft rock shores. Coast. Eng. 52 (6), 535–563.
Westoby, M., Brasington, J., Glasser, N., Hambrey, M., Reynolds, J., 2012. 'Structure-from-Motion' photogrammetry: a low-cost, effective tool for geoscience applications. Geomorphology 179, 300–314.
Wolters, G., Müller, G., 2008. Effect of cliff shape on internal stresses and rock slope stability. J. Coast. Res. 43–50.
Young, A., Flick, R., O'Reilly, W., Chadwick, D., Crampton, W., Helly, J., 2014. Estimating cliff retreat in southern California considering sea level rise using a sand balance approach. Mar. Geol. 348, 15–26.