

Modelling streambank erosion potential using maximum entropy in a central Appalachian watershed

JONATHAN PITCHFORD¹, MICHAEL STRAGER¹, ADAM RILEY¹,
LIANSHIN LIN² & JAMES ANDERSON¹

¹ West Virginia University, Division of Forestry and Natural Resources, PO Box 6125, Morgantown, West Virginia 26506-6125, USA

pitchfordj@yahoo.com

² West Virginia University, Department of Civil and Environmental Engineering, PO Box 6103, Morgantown, West Virginia 26506-6125 USA

Abstract We used maximum entropy to model streambank erosion potential (SEP) in a central Appalachian watershed to help prioritize sites for management. Model development included measuring erosion rates, application of a quantitative approach to locate Target Eroding Areas (TEAs), and creation of maps of boundary conditions. We successfully constructed a probability distribution of TEAs using the program Maxent. All model evaluation procedures indicated that the model was an excellent predictor, and that the major environmental variables controlling these processes were streambank slope, soil characteristics, bank position, and underlying geology. A classification scheme with low, moderate, and high levels of SEP derived from logistic model output was able to differentiate sites with low erosion potential from sites with moderate and high erosion potential. A major application of this type of modelling framework is to address uncertainty in stream restoration planning, ultimately helping to bridge the gap between restoration science and practice.

Key words stream restoration; maximum entropy; LiDAR; streambank erosion

INTRODUCTION

A growing number of scientists agree that where we are doing restoration and the scale at which projects are implemented, are critical for effective restoration (Wohl *et al.*, 2005). Predictive models and assessment tools that are currently used in restoration planning vary greatly in parameter selection and precision, which has implications for the scale and applications for which they are relevant (Merritt *et al.*, 2003). While several of these models have been useful for watershed management (Rosgen, 2001; Simon *et al.*, 2003), a niche remains for process models that provide a balance between high resolution prediction and broad-scale applicability. Objective decision support tools that incorporate geographic information system (GIS) and probability modelling could increase efficiency of restoration site selection and facilitate development of watershed-scale restoration plans (Wohl *et al.*, 2005) by elucidating relative streambank erosion potential (SEP) within the context of the watershed. Recent improvements and availability of remote sensing data (Goetz, 2006) used in conjunction with Bayesian reasoning has the potential to improve the resolution and precision of SEP prediction over large spatial extents (Regmi *et al.*, 2010).

We applied maximum entropy, a general purpose, machine learning method that enables prediction from incomplete information (Phillips *et al.*, 2006), to estimate the spatial distribution of target eroding areas (TEAs) undergoing excessive streambank erosion. We then used this probability distribution to create a classification scheme for streambank erosion potential (SEP). We believe this approach has great potential for enhancing watershed management by helping identify sites with the greatest restoration potential, which is critical for long-term success (Wohl *et al.*, 2005).

STUDY AREA AND METHODS

A model of SEP was constructed for a portion of the Cacapon River watershed within the larger Potomac River basin. The watershed drains about 2320 km² within Hardy, Hampshire, and Morgan counties, West Virginia. Climate is considered humid continental, characterized by hot summers, cold winters, and average annual precipitation near 900 mm. Streams in the study area typically flow through wide, slightly entrenched, shallow channels (Pitchford, 2012). We predicted SEP for 113 km of 1st–3rd order streams with median daily flowrates of 2–31 m³/s in the mainstem and 0.04–1.2 m³/s in a representative tributary during the study period (available from the USGS at

www.waterdata.usgs.gov). Elevation within the study area ranged within 210–423 m and the underlying geology consisted of alluvium (47%), shale (29%), sandstone (17%), and limestone (6%) (West Virginia Geological and Economic Survey, 2011).

Erosion rates were monitored at a total of 151 sites distributed among 30 stream reaches using erosion pin and streambank profile surveys (Hupp *et al.*, 2009) during 2010–2011. Streambank migration rates were quantified using repeated measurements of 122 cm long, 0.95 cm diameter reinforcing rods to calculate an average migration rate for each site in m/year (Hupp *et al.*, 2009). Also, repeated streambank profile surveys were conducted by measuring the horizontal distance from a level survey rod to the face of the streambank at 15 cm vertical increments to calculate a rate of change in sediment storage for each site. To determine which survey locations represent TEAs, we used cluster and outlier analysis (Anselin Local Moran's I) within the program ESRI® ArcMap™ 10.0 to determine locations of significant clustering of high erosion rates. Sites that had statistically significant clustering of high erosion rates ($Z > 1.65$; $\alpha < 0.1$) using either survey method were considered to be TEAs.

Airborne Light Detection and Ranging (LiDAR) was flown over the study area in April 2010 by the West Virginia University Natural Resource Analysis Center (WVUNRAC). Data were captured at an altitude of 1676 m and a speed of 135 knots using an Optech Inc. (Ontario, Canada) ALTM3100 with a vertical accuracy of 15 cm. These data were post-processed to create models for bare ground and vegetation within the study area. LiDAR and other available data were used to create environmental layers to represent features associated with streambank erosion (Table 1).

Table 1 Environmental variables used to model streambank erosion potential (SEP) in the Cacapon River Watershed, West Virginia.

Variable	Source	Significance	Min	Max	Mean (SE)
Geology	WVGES	Geological characteristics	N/A	N/A	N/A
Elevation	Bare earth DEM	Position in the watershed	209	423	292 (0.06)
Slope	Calculated	Steepness	0	81	9.6 (0.01)
Aspect	Calculated	Orientation, exposure	0	360	181 (0.11)
Curvature	Calculated	Steepness, concavity, convexity	−910	669	−0.2 (0.02)
Profile curvature	Calculated	Steepness, concavity, convexity	−395	481	0.2 (0.01)
Plan curvature	Calculated	Steepness, concavity, convexity	−514	378	0.0 (0.01)
Solar radiation	Calculated	Sub-aerial processes	93 189	1 423 170	1 229 151 (117)
Flow accumulation	Calculated	Run-off velocity, potential energy	0	24 641 804	119 192 (1 354)
Moisture index	Calculated	Soil water content	−4	17	0.0 (0.00)
Bank stress index	Calculated	Shear stress	N/A	N/A	N/A
Vegetation height	Vegetation DEM	Surcharge, buffer characteristics	−11	77	7.8 (0.01)
Soil type	SSURGO	Soil characteristics	N/A	N/A	N/A
Soil erodibility	SSURGO	Soil shear strength	0	0.4	0.2 (0.00)

The computer program, Maxent, version 3.3.2, was used to model SEP by estimating the unknown distribution (π) over the set of pixels in the study area. Maxent assigned a probability of occurrence to each point (x), that is approximated by solving for the entropy of $\hat{\pi}$ using the equation:

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x) \quad (1)$$

where \ln is the natural logarithm, and $\hat{\pi}$ is a positive value representing the probability of occurrence for the target phenomena that sums to one over the pixels in the study extent.

Thirty replicate bootstrap runs were conducted using 25% of the training sites that represent TEAs. Evaluation of model performance included a threshold-dependent, one-tailed binomial test on model omission and predicted area to determine if the maximum entropy distribution was predicting better than random. A threshold independent, area under curve (AUC) analysis was also used, where a value of < 0.5 indicates the model predicts no better than random, 0.5 – 0.7 indicates fair predictive capacity, 0.7 – 0.9 indicates a good predictive capacity, and values > 0.9 are indicative

of an excellent model (Phillips *et al.*, 2006). The influence of environmental variables on the distribution of TEAs was also generated by Maxent to help assess the influence of each environmental variable on the prediction.

A final map was created from the logistic model output to represent three levels of SEP (i.e. low, moderate, and high). We conducted a one-way Analysis of Variance (ANOVA) to test the strength of our classification scheme, using normally distributed migration rates (m/year) as the dependent variable. A significant ANOVA was followed with a Tukey's Honest Significant Difference (HSD) *post hoc* test to compare migration rates between low, moderate, and high SEP. Significance for all tests was set at the $\alpha = 0.05$ level.

RESULTS

Migration rates ranged from -0.11 to 0.95 m/year with an average migration rate of 0.24 m/year (SE = 0.02). Net change in sediment storage ranged from a net loss of 3.04 m²/year to a net gain of 0.80 m²/year with an average net loss of 0.39 m²/year (SE = 0.06). A total of 29 TEAs were identified from the 151 monitored locations. Twenty-five TEAs were detected based on migration rates of 0.53 – 0.93 m/year, and six TEAs were detected based on net sediment losses of 1.2 – 2.29 m²/year.

Nine of the original 14 environmental variables contributed unique information to the model and were used to create the final model of SEP. The average training AUC value for 30 model runs was 0.994 (SE = 0.0004), which indicated that the model had excellent predictive capacity. The binomial omission test was significant ($P < 0.01$, one tailed) for all data partitions at all selected threshold values indicating that the model predicted much better than random. The average logistic threshold for the minimum training presence (MTP) for all model runs was 0.209 (SE = 0.02). All logistic threshold values greater than the MTP were considered to have moderate or high SEP, which included 3.1% of the study extent. The most important environmental variables in the model were slope (32.7%), soil type (29.2%), bank stress index (20.6%), and underlying geology (8.7%) (Table 2).

Table 2 Average percent contribution and permutation importance values for each predictor variable in a maximum entropy model of stream bank erosion potential (SEP).

Variable	Percent contribution (%)	Permutation importance ^a (%)
Slope	32.7	48.6
Soil type	29.2	25.9
Bank stress index	20.6	9.1
Geology	8.7	5.2
Moisture index	3.6	7.2
Vegetation height	2.6	0.9
Elevation	1.4	2.6
Aspect	1.2	0.4
Plan Curvature	0.1	0

^a Permutation importance values for each variable are determined by randomly permuting the variable values among the training points and quantifying the ensuing decrease in training AUC.

The logistic probability of a TEA increased with increasing slope up to approximately 25° and then declined as slope increased until an asymptote was reached just above a logistic probability of 50% (Fig. 1(a)). Potomac soils were associated with the highest probability followed closely by Fluvaquents and Philo-2 soils compared to other soil types (Table 3; Fig. 1(b)). Areas along the outside of meander bends had the highest probability of being a TEA followed by the inside of meander bends, and other levels of bank stress having similar probability (Fig. 1(c)).

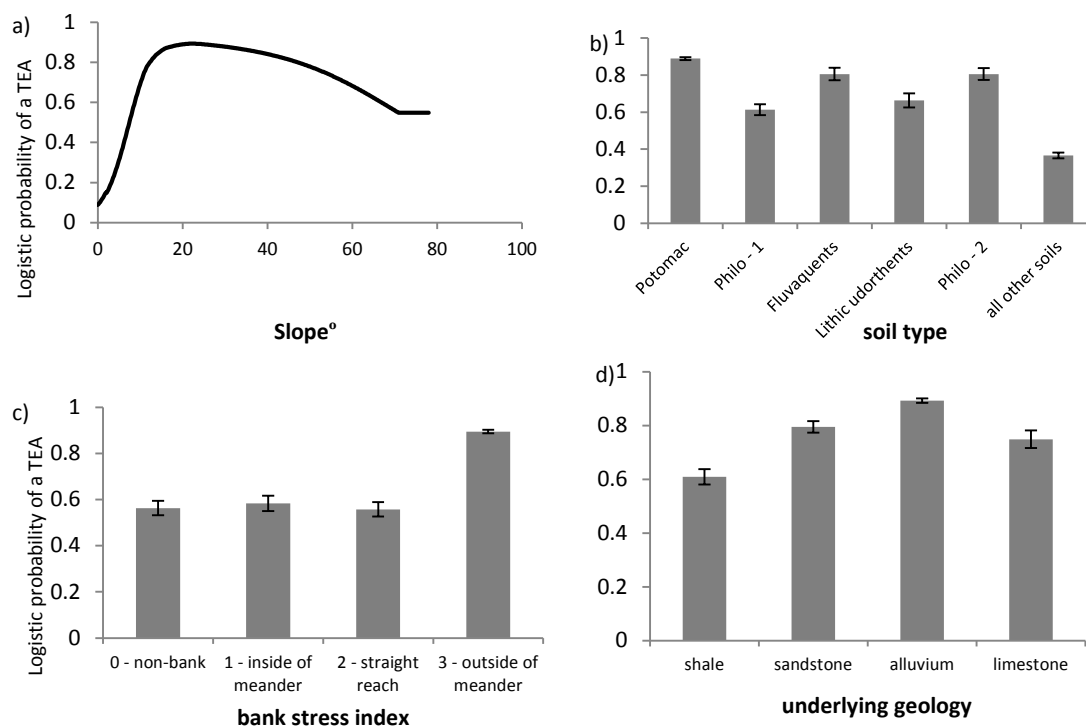


Fig. 1 The influence of: (a) slope ($^{\circ}$), (b) soil type, (c) bank stress index, and (d) underlying geology on streambank erosion potential (SEP) in the Cacapon River watershed.

Table 3 Percent sand, bulk density, soil erodibility for highly erodible soil types in the Cacapon River Watershed. Percent sand and bulk density are estimated ranges from 0–150 cm soil depth. Soil erodibility is an average value over 0–150 cm.

Soil type	Percent sand (%)	Bulk density	Soil erodibility (K_f)
Potomac	43 – 100	1.2 – 1.6	0.24
Philo – 1	0 – 95	1.2 – 1.4	0.32
Fluvaquents	0 – 60	1.0 – 1.5	0.39
Lithic udorthents	0 – 50	1.2 – 1.4	0.32
Philo – 2	0 – 95	1.2 – 1.4	0.29

With regard to underlying geology, areas comprised of alluvium had the highest probability of being a TEA with areas containing sandstone, limestone, and shale units exhibiting respective decreases in probability (Fig. 1(d)).

Our classification scheme built from logistic model output shows that 96.9% (8.5 km²) of the study extent was below the MTP, and therefore had low SEP (Fig. 2). Areas with moderate and high SEP made up 2.7% (0.24 km²) and 0.3% (0.03 km²) of the study extent, respectively. An ANOVA revealed that our classification scheme was a reliable predictor of streambank migration rate ($F_{1,149} = 33.2$; $P < 0.001$), as sites with low SEP, which averaged 0.22 m/year (SE = 0.02) were different from sites with moderate SEP with an average of 0.41 m/year (SE = 0.03) ($P < 0.001$), and from sites with high SEP, which averaged 0.45 m/year (SE = 0.04) ($P < 0.001$).

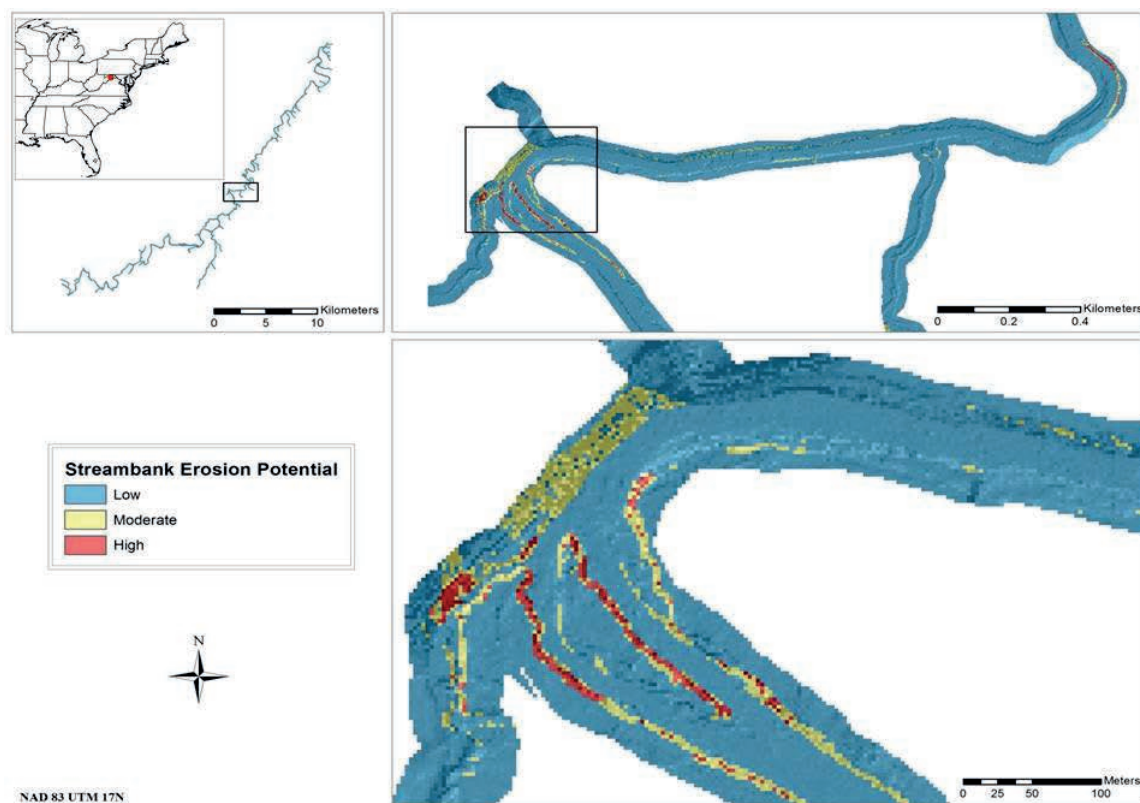


Fig. 2 Logistic output from a maximum entropy model of streambank erosion potential (SEP) in a portion of the Cacapon River Watershed, West Virginia.

CONCLUSIONS AND PERSPECTIVES

All model validation procedures indicated that our model performed well and our classification scheme was useful for predicting SEP. Thus, we believe this approach could be applied in other watersheds to enhance management by providing a high resolution prediction over large spatial extents. The most important predictor was slope where bank slopes of 25° had the highest probability of being a TEA. Steeper slopes are common in the watershed, but are typically composed of shale and thus have greatly reduced erosion rates compared to alluvial reaches. Soil type was also important as Fluvaquents and Philo soils, which contain as much as 85% sand in deeper horizons (i.e. 1–2 m) and Potomac soils, which contain as much as 100% sand (USDA, 2011), were associated with high SEP. Deeper soil horizons are often exposed in incised channels where soils with high sand content are very susceptible to fluvial erosion (Micheli & Kirchner, 2002; Simon *et al.*, 2008; Pitchford, 2012). The outside of meander bends had higher SEP compared to other levels of bank stress. This was not surprising as these areas are exposed to the highest amount of shear stress (Bloom, 1998). With regard to underlying geology, alluvium had the highest SEP. This was also not surprising as alluvium is previously eroded material (Bloom, 1998). Overall, the influence of boundary conditions was in agreement with other studies that have shown streambank slope, soil characteristics, bank position, and underlying geology are important predictors of SEP (Simon *et al.*, 2003).

Our classification scheme was effective for differentiating sites with low SEP from sites with moderate and high SEP, but could not distinguish between moderate and high levels of SEP. Although variability in streambank migration in these classes overlapped, a larger sample size among sites with high SEP would improve the ability to detect a difference. Overall, our results show that this approach has utility for gauging relative stability at the watershed (50–500 km²), segment (100–10 000 m²), and reach (10–1000 m²) scale and could help prevent unnecessary construction in areas that are relatively stable, yet may appear to be degrading. Such areas can

become a liability following restoration activity that results in reductions in flood plain roughness, which can cause bank failure (Smith & Prestegard, 2005; Pitchford, 2012).

We created only three levels of SEP from our model, but we could have easily created more categories of SEP, or generated a continuous prediction to enhance relative comparisons within the watershed. This could be very insightful for prioritizing sites for management and could help avoid attempts to stabilize streambanks with low probability of success. For example, the model can help to differentiate between areas with clay soils positioned along straight reaches (lower SEP), which have higher probability of restoration success compared to areas with sandy soils on the outside of meander bends (high SEP). Although an area with high SEP may erode at higher rates, it may be too dynamic to attempt streambank stabilization. Other applications for the model include greater understanding of conditions associated with stable sites within the watershed, which could help to inform restoration design similar to the reference reach approach used in Natural Channel Design (NCD) (Rosgen, 1998).

Overall, we believe that this maximum entropy model of SEP is a great example of an assessment tool that could enhance watershed planning by helping to prioritize sites for management, assess the relative importance of boundary conditions, and identify characteristics associated with stable sites within the watershed. This type of process model is critical for bridging the gap between restoration science and practice, which will ultimately improve the success of watershed management initiatives worldwide.

Acknowledgements The authors thank the National Fish and Wildlife Foundation, Canaan Valley Institute, the Chesapeake Bay Trust, the National Oceanic and Atmospheric Administration, FishAmerica Foundation, Cacapon and Lost Rivers Land Trust, the West Virginia University (WVU) Environmental Research Center, and the WVU NRAC for funding and logistic support. This is scientific article number 3216 of the WVU Agriculture and Forestry Experiment Station.

REFERENCES

- Bloom, A.L. (1998) *Geomorphology: A Systematic Analysis of Late Cenozoic Landforms*. Prentice Hall.
- Goetz, S.J. (2006) Remote sensing of riparian buffers: past progress and future prospects. *J. Am. Water Resour. As.* 42, 133–143.
- Hupp, C.R., *et al.* (2009) Bank erosion along the dam-regulated lower Roanoke River, North Carolina. *Geol. S. Am.* 451, 97–108.
- Merritt, W.S., Letcher, R.A. & Jakeman, A.J. (2003) A review of erosion and sediment transport models. *Environ. Modell. Softw.* 18, 761–799.
- Micheli, E.R. & Kirchner, J.W. (2002) Effects of wet meadow riparian vegetation on streambank erosion two: Measurements of vegetated bank strength and consequences for failure mechanics. *Earth Surf. Proc. Land.* 27, 687–697.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Pitchford, J.L. (2012) Stream restoration: site selection and project evaluation. Ph.D. Dissertation. West Virginia University, Morgantown, West Virginia, USA.
- Regmi, N.R., Giadino, J.R. & Vitek, J.D. (2010) Modeling susceptibility to landslides using the weight of evidence approach: western Colorado, USA. *Geomorphology* 115, 172–187.
- Rosgen, D.L. (1998) The Reference Reach – A Blueprint for Natural Channel Design. In: *Restoration of Wetlands and Rivers*, Proceedings of the American Society of Civil Engineers, Denver, Colorado.
- Rosgen, D. (2001) Practical method of computing streambank erosion rate. Proceedings of the Seventh Federal Interagency Sedimentation Conference, Reno, Nevada.
- Simon, A., Collison, A.J.C. & Layzell, A. (2003) Incorporating bank-toe erosion by hydraulic shear into the ARS bank stability model: Missouri River, eastern Montana. Proceedings of the First Interagency Conference on Research in the Watersheds, USDA ARS, Arizona 70–76.
- Simon, A., *et al.* (2008) Reply to discussion by Dave Rosgen “Critical evaluation of how the Rosgen classification and associated ‘Natural Channel Design’ methods fail to integrate and quantify fluvial processes and channel responses.” *J. Am. Water Resour. As.* 44, 782–792.
- Smith, S.M. & Prestegard, K.L. (2005) Hydraulic performance of a morphology-based stream channel design. *Water Resour. Res.* 41, 1–17.
- USDA (2011) Soil Survey Geographic Database (SSURGO) for West Virginia.
- West Virginia Geological and Economic Survey (2011) Statewide geologic map obtained from Mont Chateau Research Center, West Virginia.
- Wohl, E., *et al.* (2005) River restoration. *Water Resour. Res.* 41, 1–12.