

Slope Failure Probabilistic Mapping with Random Forests

Your results should be delivered as an HTML webpage generated using R Markdown. Make sure to include all code and results. Provide text to describe your methods and results. This should read like the Methods and Results sections of a paper.

Grading Criteria

- Correctness and completeness of code (16 Points)
- Description of process and results (12 Points)
- Webpage formatting (6 Points)
- Map output (6 Points)

In this assignment you will use machine learning to create a probabilistic prediction of slope failure occurrence. These data were generated for a mapping project undertaken by the West Virginia GIS Technical Center to predict the likelihood of slope failure occurrence for the entire Valley and Ridge section of West Virginia. Slope failure locations were mapped by interpretation of light detection and ranging (LiDAR) data, and pseudo-absence data were generated as random points not within 30 meters of the mapped slope failures. A variety of predictor variables were produced relating to terrain conditions, lithology, soil characteristics, and distance from roads and streams as described in the provided tables.

Use the **randomForest** package to create a random forest model to predict the likelihood of slope failure occurrence. Create models using all predictor variables, just the terrain variables, just lithology and soil variables, just distance variables, and all non-terrain variables. Compare the models using ROC curves and the AUC statistic as implemented in the **pROC** package. Predict to the raster stack (**stack.img**) to create a spatial prediction, and display the result using **tmap**.

Preparation

- Read in the table of samples (**lsm_data.csv**) and the raster stack of predictor variables (**stack.tif**).
- Define the “steve”, “dspm”, and “drain” columns as factors.
- Extract 250 slope failure (“slopeD”) and 250 pseudo-absence (“not”) samples for validation.
- Extract the remaining 1,500 slope failure (“slopeD”) and 1,500 pseudo-absence (“not”) samples for training. These should not overlap with the validation samples.

Training

- Train separate models for each of the feature spaces. For all models the dependent variable will be the “class” column (the first column).
 - All: columns 2 through 44
 - Terrain Only: columns 2 through 33
 - Lithology and Soil: columns 34 through 44
 - Distance Only: columns 34 through 41
 - All Non-Terrain: columns 34 through 44

Validation

- Predict the validation data with all the models.
- Generate ROC curves and the AUC metric for each model using the validation results.

- Create a ROC plot containing all ROC curves for comparison.
- Write a paragraph comparing the results for the different feature spaces.

Spatial Prediction

- Change the raster grid band names to match the names in the table. The names should match and are in the same order as columns 2 through the final column in the original table (**lsm_data2.csv**)
- Use the model that was trained with all predictor variables to predict to the raster stack. Note that this will take some time.
- Read the result back in and display it using **tmap**. Make sure to provide a legend, title, and use an appropriate color palette.

Table 1. Terrain predictor variables.

Variable	Abbreviation	Description	Window Sizes (cells)
Slope	slp	Gradient or rate of maximum change in Z as degrees of rise	1
Mean Slope	slpmn	Slope average over a local window	7, 11, 21
Linear Aspect	asp_lin	Transform of topographic aspect to linear variable	1
Profile Curvature	proc	Curvature parallel to direction of maximum slope	7, 11, 21
Plan Curvature	planc	Curvature perpendicular to direction of maximum slope	7, 11, 21
Longitudinal Curvature	longc	Profile curvature intersecting with the plane defined by the surface normal and maximum gradient direction	7, 11, 21
Cross-Sectional Curvature	crossc	Tangential curvature intersecting with the plane defined by the surface normal and a tangent to the contour - perpendicular to maximum gradient direction	7, 11, 21
Slope Position	sp	Z – Mean Z	7, 11, 21
Topographic Roughness	rph	Square root of standard deviation of slope in local window	7, 11, 21
Topographic Dissection	diss	$\frac{Z - \text{Min } Z}{\frac{\text{Max } Z - \text{Min } Z}{\text{Cell Area}}}$	7, 11, 21
Surface Area Ratio	sar	$\frac{\text{Cosine}(\text{slope} * \pi * 180)}{\text{Mean } Z - \text{Min } Z}$	1
Surface Relief Ratio	ssr	$\frac{\text{Mean } Z - \text{Min } Z}{\text{Max } Z - \text{Min } Z}$	7, 11, 21
Site Exposure Index	sei	Measure of exposure based on slope and aspect	1
Heat Load Index	hli	Measure of solar insolation based on slope, aspect, and latitude	1

Table 2. Additional predictor variables.

Variable	Abbreviation	Description
Distance to Roads	us_dist, state_dist, local_dist	Euclidean distance to nearest US, state, and local road
Cost Distance to US Roads (US, state, and local)	us_cost, state_cost, local_cost	Euclidean distance to nearest US, state, and local road weighted by slope
Distance from Streams	strm_dist	Distance from mapped streams
Cost Distance from Streams	strm_cost	Distance from
Geomorphic Presentation	Steve	Classification of rock formations based on geomorphic presentation
Dominant Soil Parent Material	dspm	Dominant parent material of soil
Soil Drainage Class	drain	Drainage class of soil