Using Logistic Regression To Predict the Probability of Debris Flows Occurring in Areas Recently Burned By Wildland Fires

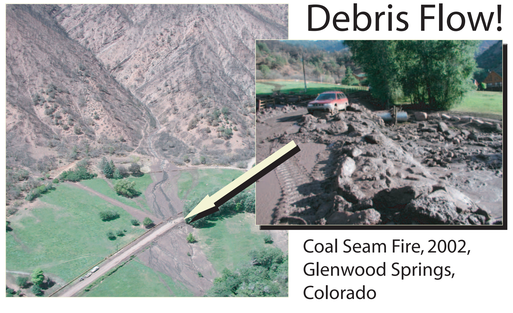
By Michael G. Rupert, Susan H. Cannon, and Joseph E. Gartner

Open-File Report OF 03-500

Missionary Ridge Fire, 2002, Durango, Colorado

Abstract

Logistic regression was used to predict the probability of debris flows occurring in areas recently burned by wildland fires. Multiple logistic regression is conceptually similar to multiple linear regression because statistical relations between one dependent variable and several independent variables are evaluated. In logistic regression, however, the dependent variable is transformed to a binary variable (debris flow did or did not occur), and the actual probability of the debris flow occurring is statistically modeled. Data from 399 basins located within 15 wildland fires that burned during 2000-2002 in Colorado, Idaho, Montana, and New Mexico were evaluated. More than 35 independent variables describing the burn severity, geology, land surface gradient, rainfall, and soil properties were evaluated. The models were developed as follows: (1) Basins that did and did not produce debris flows were delineated from National Elevation Data using a Geographic Information System (GIS). (2) Data describing the burn severity, geology, land surface gradient, rainfall, and soil properties were determined for each basin. These data were then downloaded to a statistics software package for analysis using logistic regression. (3) Relations between the occurrence/non-occurrence of debris flows and burn severity, geology, land surface gradient, rainfall, and soil properties were evaluated and several preliminary multivariate logistic regression models were constructed. All possible combinations of independent variables were evaluated to determine which combination produced the most effective model. The multivariate model that best predicted the occurrence of debris flows was selected. (4) The multivariate logistic regression model was entered into a GIS, and a map showing the probability of debris flows was constructed. The most effective model incorporates the percentage of each basin with slope greater than 30 percent, percentage of land burned at medium and high burn severity in each basin, particle size sorting, average storm intensity (millimeters per hour), soil organic matter content, soil permeability, and soil drainage. The results of this study demonstrate that logistic regression is a valuable tool for predicting the probability of debris flows occurring in recently-burned landscapes.



Coal Seam Fire, 2002, Glenwood Springs, Colorado

Maps predicting the probability of debris flows were developed in four steps

(1) The first step was to delineate basins that did and did not produce debris flows using National Elevation Data (NED). A total of 399 basins located within 15 wildland fires in 4 western states were delineated. As an example, basins from the Missionary Ridge Fire in Colorado are shown in figure 1.

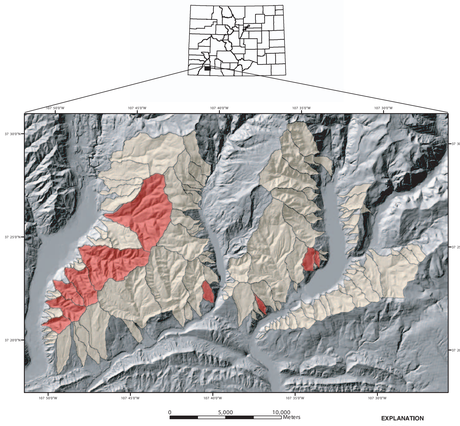


Figure 1. Basins with and without debris flows, Missionary Ridge Fire, 2002, near Durango, Colorado.

(2) The second step was to determine the burn severity (fig. 2), geology, land surface gradient, rainfall, and soil properties for each of the 15 wildland fires.

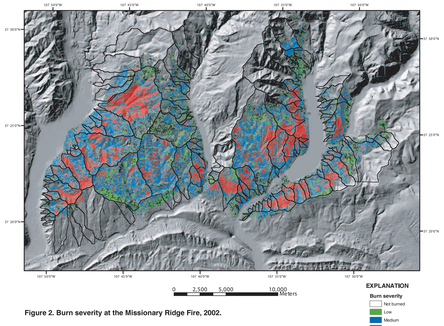


Figure 2. Burn severity at the Missionary Ridge Fire, 2002.

(3) The third step was to observe statistical correlations of data from all 15 wildland fires and build the logistic regression models (table 1) by evaluating all possible combinations of variables.

- **Univariate correlations can be a good initial indicator of which independent variables might be significant in the multivariate models. Univariate correlations are usually observed first to explore the data (table 1).**
- **Multivariate models were constructed by evaluating every possible combination of variables shown in table 1. Several logistic regression models were constructed using different combinations of variables, and then the most effective model was selected.**

Table 1. Summary of univariate and multivariate correlations between debris flow occurrence and independent variables describing burn severity, geology, land surface gradient, rainfall, soil properties, and variable interactions using data from 15 wildland fires in Colorado, Idaho, Montana, and New Mexico, 2000-2002.

[GE, greater than or equal to; mm, millimeters; hr, hour; >, greater than; < less than; X, denotes statistically significant variable in multivariate correlations]

VARIABLES EVALUATED BY THIS STUDY	UNIVARIATE CORRELATIONS		MULTIVARIATE CORRELATIONS (in decreasing order of significance. Significant variables denoted with an X)							
			Model Number	M1	M2	M3	M4	M5	M6	M7
			Spearman's Rho (a)	0.531	0.529	0.527	0.522	0.488	0.480	0.479
			Wilcoxon Test (b)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			McFadden's Rho (c)	0.367	0.377	0.367	0.311	0.298	0.280	0.293
BURN SEVERITY	Spearman's Rho (a)	McFadden's Rho (c)								
Percent of Total Area In Each Basin That Was Burned at Low Severity	-0.352	0.000								
Percent of Total Burned Area In Each Basin That Was Burned at Low Severity	-0.400	0.123								
Percent of Total Area In Each Basin That Was Burned at Medium Severity	0.185	0.026								
Percent of Total Burned Area In Each Basin That Was Burned at Medium Severity	0.089	0.003								
Percent of Total Area In Each Basin That Was Burned at High Severity	0.193	0.014								
Percent of Total Burned Area In Each Basin That Was Burned at High Severity	0.166	0.051								
Percent of Total Area In Each Basin That Was Burned at Medium Plus High Severity	**0.416**	0.043					X			
Percent of Total Burned Area In Each Basin That Was Burned at Medium Plus High Severity	0.396	**0.103**		X	X	X		X	X	X
Percent of Total Area Burned at All Severities	0.065	0.022								
GEOLOGY										
Sedimentary	-0.218	0.000								
Granitic	0.022	0.000								
Metamorphic	0.116	0.017								
Volcanic	0.225	0.042								
LAND SURFACE GRADIENT										
Average Gradient (percent)	0.054	0.013								
Gradient Times Total Area	-0.030	0.005								
Gradient Divided by Total Area	0.024	0.005					X			X
Percent Total Basin Area with Slopes GE 30 percent	-0.058	0.007		X	X	X		X	X	
Percent Total Basin Area with Slopes GE 50 percent	0.069	0.023								
Average aspect (degrees from north)	-0.062	0.000								
Ruggedness (change in basin elevation/square-root of area)	**0.281**	**0.035**								X
RAINFALL										
Storm Total (mm)	0.099	0.021					X			
Storm Duration (hr)	0.261	0.000								
Average Storm Intensity (mm/hr)	0.057	0.011		X	X	X		X	X	
Peak 10-minute Rainfall Intensity for a Given Storm (mm/hr)	**0.317**	0.018								
Peak 15-minute Rainfall Intensity for a Given Storm (mm/hr)	0.222	**0.038**								X
Peak 30-minute Rainfall Intensity for a Given Storm (mm/hr)	0.235	0.022					X			
Peak 60-minute Rainfall Intensity for a Given Storm (mm/hr)	0.248	0.035								
SOILS PROPERTIES										
Median Particle Size (phi units)	**0.386**	**0.024**								
Mean Particle Size (phi units)	**-0.176**	**0.029**								
Particle Size Sorting	**-0.196**	**0.050**		X					X	X
Particle Size Skewness	**0.258**	0.004								
Available Water Capacity (inches per inch)	0.046	0.006					X			
Clay Content (percent of material < 2mm)	**-0.163**	0.000			X	X	X		X	X
Soil Erodibility Factor (KFFACT, 1 = easily eroded, 0.01 = non erosive)	-0.074	**0.026**								
Organic Matter (percent, by weight)	**-0.172**	0.006		X	X	X	X	X	X	
Permeability (inches per hr)	-0.118	0.004		X	X	X	X	X	X	X
Soil Thickness (inches)	**-0.147**	**0.020**			X	X	X		X	
Hydrologic Group (1 though 4, with 1 having highest infiltration)	0.111	**0.022**			X	X	X	X		
Drainage (1 through 7, with 1 having highest drainage)	**-0.198**	0.006		X	X	X	X	X	X	X
Slope (in percent)	-0.066	0.004					X			
Liquid Limit (percent moisture by weight)	0.046	0.004			X	X		X		
Hydric Capacity (used to identify wetlands, proportion of map unit with saturated (anaerobic) soils)	**-0.190**	0.002								
Annual Flood Frequency (1 to 4, 1 > 50 percent, 2 = 5 to 50 percent, 3 = 0 to 5 percent, 4 = none)	-0.005	0.003			X		X	X		
VARIABLE INTERACTIONS										
Percent of Total Burned Area In Each Basin That Was Burned at Medium Plus High Severity * Soil Organic Matter	n/a	n/a		X	X	X				
Sorting*Drainage	n/a	n/a		X						
Percent Slope GE 30%.* Average Storm Intensity	n/a	n/a		X	X	X				
Soil Permeability * Soil Drainage	n/a	n/a			X	X				

(a) Spearman's *rho* is a nonparametric rank-order test, similar to the student's t-test. Variables and models with no correlation with debris flows have spearman's *rho* numbers near zero, and variables with perfect correlations have numbers near plus or minus one.

(b) The Wilcoxon test is a nonparametric test to determine if there is a significant difference between areas with and without debris flows. As the effectiveness of the models improves, the Wilcoxon p-values approach zero.

(c) McFadden's *rho* is calculated by logistic regression, and conceptually is similar to an *r*-squared in linear regression. A McFadden's *rho* value of zero means there is no correlation; values between 0.2 and 0.4 denote significant correlation.

(4) The fourth step was to enter the most effective model into the GIS, and construct maps predicting the probability of debris flows occurring. As an example, the probability of debris flows occurring in the Missionary Ridge Fire is shown in figure 3.

$$\text{Probability of Debris Flow} = \frac{e^{(-29.693 + 2.864*SL + 10.697*BS - 9.875*SO + 0.208*SI + 5.729*OM - 0.957*P + 9.351*D - 8.335*BS*OM + 4.669*SO*D - 0.174*SL*SI)}}{1 + e^{(-29.693 + 2.864*SL + 10.697*BS - 9.875*SO + 0.208*SI + 5.729*OM - 0.957*P + 9.351*D - 8.335*BS*OM + 4.669*SO*D - 0.174*SL*SI)}}$$

Where:

SL = Percent of basin with slope greater than or equal to 30 percent
BS = Burn severity category
SO = Partical size sorting, in Phi units
SI = Average storm intensity, in millimeters per hour
OM = Soil organic matter, in percent by weight
P = Soil permeability, in inches per hour
D = Soil drainage category

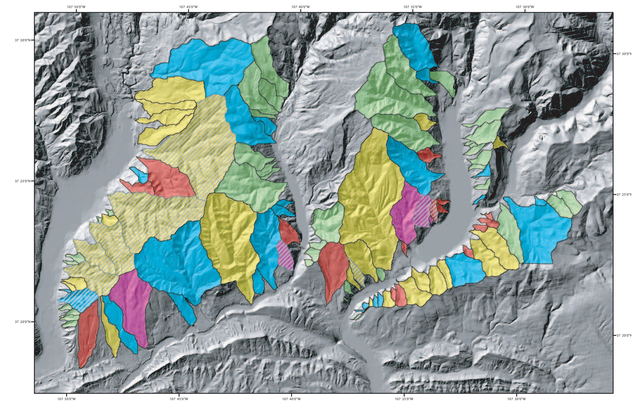


Figure 3. Probability of a debris flow occurring in response to a 25-year rain event (33 mm in one hour), Missionary Ridge Fire, 2002, using model M1.

References

Hosmer, D.W., and Lemeshow, S., 1989, Applied logistic regression: New York, John Wiley & Sons, Inc., 307 p.

Schwarz, G.E. and Alexander, R.B. 1995, State Soil Geographic (STATSGO) Data base for the conterminous United States: U.S. Geological Survey Open-File Report 95-449, http://water.usgs.gov/lookup/getspatial?ussoils

Acknowledgments

**Montana Bureau of Mines and Geology,
U.S.D.A. Forest Service,
U.S. Geological Survey**