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

Abstract

Logistic regression was used to predict the probability of debris flows occurring in areas recently burned by wildland fires. Multivariate 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 in a binary variable (either 0 or 1). When a fire occurs, the dependent variable is 1, and when a fire does not occur, the dependent variable is 0. The area burned in 15 wildfires that burned during 2000-2002 in Colorado, Idaho, Montana, and New Mexico was evaluated. More than 25 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 using a Geographic Information System (GIS), and the burn severity, geology, land surface gradient, rainfall, and soil properties were determined for each basin. These data were then imported to a standard software package for analysis using logistic regression. (2) Relationships between burn severity and multiple logistic regression models with dependent variables were evaluated to determine those relationships that best predicted the occurrence of debris flows. (3) The third step was to summarize the relationships between variables to predict the occurrence of debris flows. (4) The fourth step was to enter the most effective model into the GIS, and construct maps predicting the probability of debris flows occurring in the Missionary Ridge Fire is shown in figure 3.

Debris Flow!

Probability of Debris Flow

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 wildfires in 4 western states were delineated. As an example, basins from the Missionary Ridge Fire in Colorado are shown in figure 1.

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 wildfires.

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

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.


Acknowledgments


References


Spearman's rho numbers near zero, and variables with perfect correlations have numbers near plus or minus one. (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

Table 1:  Summary of multivariate logistic regression analysis. Independent variables were evaluated in logistic regression models in 4 steps. Probability of debris flow was calculated using different combinations of variables, and then the most effective model was selected.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Logistic Regression</th>
<th>Probability of Debris Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burn Severity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land Surface Gradient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil Properties</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>