Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, Southern California, 2003–2006

Open-File Report 2008–1370

U.S. Department of the Interior
U.S. Geological Survey

Cover. Debris flows in areas burned by the 2007 Poomacha Fire, near San Diego, California. Photographs by George Wilkens, Pacific REMS, Inc.

Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, Southern California, 2003–2006

By Michael G. Rupert, Susan H. Cannon, Joseph E. Gartner, John A. Michael, and Dennis R. Helsel

Open-File Report 2008–1370

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2008

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Rupert, M.G., Cannon, S.H., Gartner, J.E., Michael, J.A., and Helsel, D.R., 2008, Using logistic regression to predict the probability of debris flows in areas burned by wildfires, southern California, 2003–2006: U.S. Geological Survey Open-File Report 2008–1370, 9 p.

Contents

Figures

Table

Conversion Factors

SI to Inch/Pound

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

°F=(1.8×°C)+32

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

°C=(°F-32)/1.8

Vertical coordinate information is referenced to the insert datum name (and abbreviation) here, for instance, “North American Vertical Datum of 1988 (NAVD 88)”

Horizontal coordinate information is referenced to the insert datum name (and abbreviation) here, for instance, “North American Datum of 1983 (NAD 83)”

Altitude, as used in this report, refers to distance above the vertical datum.

Abbreviations Used in This Report

GIS	geographic information system
in.	inch, inches
hr	hour, hours
km	kilometers
m	meter, meters
mm	millimeter, millimeters

Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, Southern California, 2003–2006

By Michael G. Rupert, Susan H. Cannon, Joseph E. Gartner, John A. Michael, and Dennis R. Helsel

Abstract

Logistic regression was used to develop statistical models that can be used to predict the probability of debris flows in areas recently burned by wildfires by using data from 14 wildfires that burned in southern California during 2003–2006. Twenty-eight independent variables describing the basin morphology, burn severity, rainfall, and soil properties of 306 drainage basins located within those burned areas were evaluated. The models were developed as follows: (1) Basins that did and did not produce debris flows soon after the 2003 to 2006 fires were delineated from data in the National Elevation Dataset using a geographic information system; (2) Data describing the basin morphology, burn severity, rainfall, and soil properties were compiled for each basin. These data were then input to a statistics software package for analysis using logistic regression; and (3) Relations between the occurrence or absence of debris flows and the basin morphology, burn severity, rainfall, and soil properties were evaluated, and five multivariate logistic regression models were constructed. All possible combinations of independent variables were evaluated to determine which combinations produced the most effective models, and the multivariate models that best predicted the occurrence of debris flows were identified. Percentage of high burn severity and 3-hour peak rainfall intensity were significant variables in all models. Soil organic matter content and soil clay content were significant variables in all models except Model 5. Soil slope was a significant variable in all models except Model 4. The most suitable model can be selected from these five models on the basis of the availability of independent variables in the particular area of interest and field checking of probability maps. The multivariate logistic regression models can be entered into a geographic information system, and maps showing the probability of debris flows can be constructed in recently burned areas of southern California. This study demonstrates that logistic regression is a valuable tool for developing models that predict the probability of debris flows occurring in recently burned landscapes.

Introduction

Wildfire can have immediate and profound effects on drainage basins. Consumption of the rainfall-intercepting canopy and of the soil-mantling litter and duff, intensive drying of the soil, generation of vegetative ash, combustion of soil-binding organic matter, and the formation of water-repellent soils and surface sealing of soil pores by wood ash can decrease infiltration of rainfall and substantially increase overland flow. Removal of obstructions to flow, such as vegetative cover and live or downed timber, can increase the erosive power of overland flow and accelerate detachment and removal of soil material from hillslopes. Increased runoff can also entrain and erode substantial volumes of material from stream channels and banks. Rainfall on recently burned basins produces increased runoff that commonly transports and deposits large volumes of sediment, both within and downstream from the burned area (Cannon, 2005; Cannon and others, 2008).

Debris flows are among the most hazardous consequences of rainfall on recently burned hillslopes (Cannon, 2005; Cannon and others, in press; Stevens and others, 2008). Debris flows pose a hazard distinct from other sediment-laden flows because of their unique destructive power. They can occur rapidly, with little warning, and exert great impulsive loads on objects in their paths; even small debris flows can strip vegetation, block drainageways, damage structures, and endanger human life (Iverson, 1997). For example, a storm on Christmas Day 2003 triggered debris flows from many of the steep basins burned earlier that fall in the San Bernardino Mountains of southern California (Brock and others, 2007). Sixteen people were killed by this event, and billions of dollars were spent for cleanup and repairs (Chong and others, 2004). This event clearly illustrates the need for tools to rapidly assess debris-flow hazards immediately after wildfires.

In a study of the erosional response of recently burned basins throughout the Western United States, including southern California, Cannon (2000, 2001) found that not all burned basins that receive heavy precipitation produce debris flows. Rather, many burned drainage basins respond to even heavy rainfall by sediment-laden flooding, rather than by debris flows. It became apparent that debris flows tend to occur when specific combinations of basin morphology, burn severity, soil properties, and rainfall amounts are met. To investigate relations between the occurrence of debris flow and the controlling variables, Rupert and others (2003) completed a pilot project in which models to predict the probability of debris flows in recently burned lands of the Intermountain West were developed by using logistic regression statistical modeling methods. These models were based on correlations between the occurrence of debris flows and factors such as basin morphology, burn severity, soil properties, and rainfall amounts. Rupert and others (2003) demonstrated that logistic regression is a useful tool for developing models to predict the probability of debris flows in areas recently burned by wildfires.

In response, a series of models were developed for the Intermountain West (Cannon and others, in press). Although those models worked well in the Intermountain West, the models were not as effective in southern California, probably because of differences in climate, geology, soil types, vegetation types, and fire histories. This report describes logistic regression models designed to predict the probability of debris flows specifically in areas recently (2003–2006) burned by wildfires in southern California (fig. 1).

Approach

Logistic regression was used to develop models to predict the probability of debris flows occurring in areas recently burned by wildfires in southern Caifornia. Data from 306 drainage basins located within the Blaisdell, Day, Gaviota, Gorman, Grand Prix, Harvard, Horse, Old, Piru, School, Soboba, Thurman, Topanga, and Woodhouse wildfires that occurred during 2003 through 2007 in southern California were evaluated (fig. 1). Twenty-eight independent variables describing the basin morphology, burn severity, rainfall, and soil properties were evaluated (table 1). The models were developed as follows: (1) Basins that did and did not produce debris flows soon after the 2003 to 2006 fires were delineated from data in the National Elevation Dataset (NED) using a

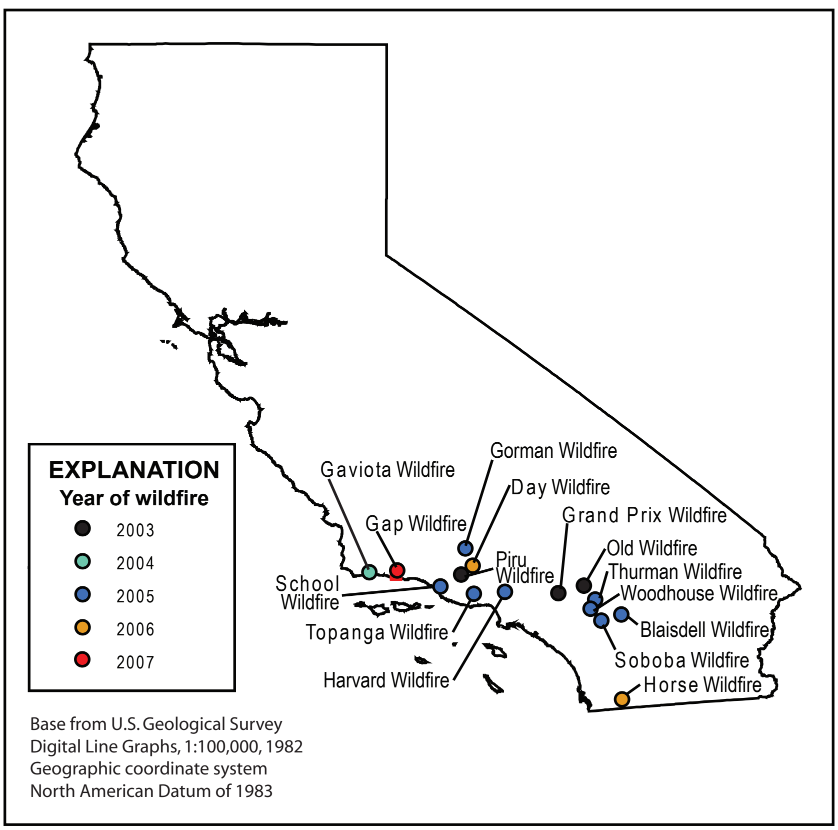


Figure 1. Areas burned by wildfires and evaluated in this study, southern California, 2003-2007.

geographic information system (GIS); (2) Measures of basin morphology, burn severity, rainfall, and soil properties were determined for each basin. These data were then input to the SYSTAT statistics software package for analysis using logistic regression; and (3) Relations between the occurrence or absence of debris flows and measures of basin morphology, burn severity, rainfall, and soil properties were evaluated and several preliminary multivariate logistic regression models were constructed. All possible combinations of independent variables were evaluated, and the statistics calculated by logistic regression were used to identify the most effective models. To illustrate the results of this modeling exercise, a GIS was used to generate a map showing the probability of debris flows in the area burned by the 2007 Gap wildfire, in southern California (fig. 1).

Logistic Regression

Logistic regression (Kleinbaum, 1994; Hosmer and Lemeshow, 2000; Helsel and Hirsch, 2002) is a statistical method that predicts the probability of an event occurring, in this case, the probability of a debris flow occurring after a wildfire. Logistic regression is conceptually similar to multiple linear regression, because relations between one dependent variable and several independent variables are evaluated. Whereas multiple linear regression returns a continuous value for the dependent variable, logistic regression returns the probability of a positive binomial outcome (in this case, debris flows did or did not occur) in the form:

$$P = e^x/(1+e^x) \quad (1)$$

where

P is the probability of debris-flow occurrence, in percent;
x is $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i$;
β_i is logistic regression coefficients;
x_i is values for the independent variables, such as rainfall amounts or soil properties; and
$_i$ is the number of variables.

Logistic regression calculates several statistical parameters that determine the predictive success of the model (Kleinbaum, 1994; Hosmer and Lemeshow, 2000). The p-values calculated for each independent variable, indicates the statistical significance that each variable has on the overall logistic regression model. A p-value of 0.1 indicates a significance level of 90 percent, and a p-value of 0.05 indicates a significance level of 95 percent. Independent variables were excluded from the models if their individual p-values were greater than 0.1; the largest p-value of an individual, independent variable in any of the final models was 0.08 (table 1). McFadden's rho-squared (SYSTAT Software, Inc., 2004) is conceptually similar to the r-squared of linear regression. McFadden's rho-squared is always between zero and one; a McFadden's rho-squared approaching 1 corresponds with more significant results. McFadden's rho-squared tends to be smaller than r-squared, so a small number does not necessarily imply a poor fit. Values between 0.20 and 0.40 indicate good results (SYSTAT Software, Inc., 2004). The percentage of correctly predicted events (sensitivity) is calculated as the correctly predicted number of events (debris flows) divided by the total number of observed or actual events (SYSTAT Software, Inc., 2004, p. II-234). The percentage of correctly predicted reference events (specificity) is calculated as the number of correctly predicted number of reference events (no debris flows) divided by the total number of observed or actual reference events (SYSTAT Software, Inc., 2004, p. II-234). To validate the models, the percentage of actual events was plotted with the predicted probability of events using a deciles-of-risk calculation, which typically partitions the sample into 10 groups (SYSTAT Software, Inc., 2004, p. II-238).

In some cases, the data distributions of the independent variables can be significantly nonlinear (skewed). The results of logistic regression modeling can be significantly improved by transforming the skewed independent variables to linear distributions using log, natural log, or square root transformations (Menard, 2002). Prior to calculation of the logistic regression models, independent variables were evaluated for skewness. The burn severity and soils data were not significantly skewed, but all of the rainfall and basin morphology data were skewed. Skewness of the rainfall and basin morphology data was significantly reduced by transforming the data using a natural log transformation; square root and log base 10 transformations were less effective. The logistic regression models were developed using the natural log transformed data, and all probability maps were constructed using GIS data that were also transformed.

Input Data

Because it was not known which factors would best predict debris-flow occurrence, this study evaluated a number of different independent variables in the logistic regression analyses (table 1).

Seven measures of basin morphology and drainage network were compiled from 30-meter (m) or 10-m NED data (U.S. Geological Survey, 1999), depending on availability. These measures were as follows:

- Drainage basin area [in square kilometers (km^2)],
- Percent slope greater than or equal to 30 percent,
- Percent slope greater than or equal to 50 percent,
- Average basin relief (in m),
- Average basin gradient (as a percentage),
- Average channel length (in km), and
- Drainage network bifurcation ratio (average ratio of the number of streams of any order to the number of streams of the next highest order (Horton, 1945) (unitless).

Table 1. Logistic regression modeling results, coefficients, and individual p-values of independent variables significantly related with the probability of debris flows in the Blaisdell, Day, Gaviota, Gorman, Grand Prix, Harvard, Horse, Old, Piru, School, Soboba, Thurman, Topanga, and Woodhouse wildfires, southern California, 2003–2006

[--, no relation observed; values not enclosed in parentheses are logistic regression coefficients; values enclosed in parentheses are individual p-values; independent variables in bold are used in at least one of the logistic regression models; ln, natural logarithm; in., inches; hr, hour; mm, millimeters; km^2, square kilometers]

Statistical measures and independent variables	Models				
Statistical measures	Model 1	Model 2	Model 3	Model 4	Model 5
McFaddens rho-squared	0.620	0.607	0.605	0.604	0.587
Sensitivity	0.756	0.746	0.745	0.750	0.745
Specificity	0.909	0.905	0.905	0.907	0.905
Logistic regression constant	–20.807	–18.598	–11.916	–18.095	–18.874
Independent variables					
Basin morphology					
Drainage basin area (km^2)	--	--	--	--	--
Percent slope greater than or equal to 30 percent	--	--	--	0.042 (0.037)	--
Percent slope greater than or equal to 50 percent	--	--	0.022 (0.080)	--	--
Average basin relief (m) (transformed to ln)	1.650 (0.006)	--	--	1.580 (0.013)	1.458 (0.005)
Average basin gradient (percent) (transformed to ln)	--	2.147 (0.062)	--	--	--
Average channel length (km) (transformed to ln)	–0.694 (0.020)	--	--	–0.705 (0.033)	–0.65 (0.017)
Drainage network bifurcation ratio (unitless) (transformed to ln)	--	--	--	--	--
Burn severity					
Percentage of the basin burned at low burn severity	--	--	--	0.023 (0.024)	--
Percentage of the basin burned at medium burn severity	--	--	--	--	--
Percentage of the basin burned at high burn severity	0.044 (0.029)	0.035 (0.057)	0.034 (0.060)	0.067 (0.006)	0.061 (0.003)
Rainfall					
Total storm rainfall (mm)	--	--	--	--	--
Total storm duration (decimal hours) (transformed to ln)	--	--	--	--	--
Maximum amount of rainfall measured over a 10–minute period divided by the time (mm/hr) (transformed to ln)	--	--	--	--	--
Maximum amount of rainfall measured over a 15–minute period divided by the time (mm/hr) (transformed to ln)	--	--	--	--	--
Maximum amount of rainfall measured over a 30–minute period divided by the time (mm/hr) (transformed to ln)	--	--	--	--	--
Maximum amount of rainfall measured over a 60–minute period divided by the time (mm/hr) (transformed to ln)	--	--	--	--	--
Maximum amount of rainfall measured over a 3–hour period divided by the time (mm/hr) (transformed to ln)	2.463 (0.000)	2.356 (0.000)	2.416 (0.000)	2.939 (0.000)	2.609 (0.000)
Maximum amount of rainfall measured over a 6–hour period divided by the time (mm/hr) (transformed to ln)	--	--	--	--	--
Soil properties					
Soil slope (percentage)	0.128 (0.002)	0.106 (0.005)	0.114 (0.003)	--	0.077 (0.008)
Soil permeability (in./hr)	--	--	--	--	0.287 (0.038)
Soil organic matter content (percent, by weight)	7.229 (0.002)	7.402 (0.002)	7.338 (0.002)	6.349 (0.003)	--
Soil erodibility factor (unitless)	--	--	--	--	--
Soil hydrologic group (1 through 4; 1 has the highest infiltration rate)	--	--	--	--	--
Soil drainage (1 through 7; 1 has the lowest drainage rate)	--	--	--	--	--
Soil clay content (percentage of material less than 2 mm)	–0.245 (0.001)	–0.288 (0.000)	–0.278 (0.000)	–0.318 (0.000)	--
Soil available water capacity (in./in.)	--	--	--	--	--
Soil thickness within the uppermost 79 inches below land surface (in.)	--	--	--	--	--
Soil liquid limit (percentage of moisture by weight)	--	--	--	--	--

Burn severity for each basin was determined by using maps of burn severity generated using the normalized burn ratio, as determined from Landsat Thematic Mapper data (Key and Benson, 2006). The maps of burn severity reflect changes in biomass immediately after a fire. Three measures of burn severity were evaluated:

- Percentage of the basin area burned at low severity,
- Percentage of the basin area burned at moderate severity, and
- Percentage of the basin area burned at high severity.

Rainfall in each basin was determined from tipping-bucket rain gages located within 2 km of each basin. The following rainfall measurements were used in the logistic regression models:

- Total storm rainfall [in millimeters (mm)],
- Storm duration (in decimal hours),
- Maximum amount of rainfall measured over a 10–minute period divided by the time [in mm/hour (hr)],
- Maximum amount of rainfall measured over a 15–minute period divided by the time (in mm/hr),
- Maximum amount of rainfall measured over a 30–minute period divided by the time (in mm/hr),
- Maximum amount of rainfall measured over a 60–minute period divided by the time (in mm/hr),
- Maximum amount of rainfall measured over a 3–hr period divided by the time (in mm/hr), and
- Maximum amount of rainfall measured over a 6–hr period divided by the time (in mm/hr).

Soils data were obtained from the State Soil Geographic (STATSGO) database (U.S. Department of Agriculture, 1991). The finer scale Soil Survey Geographic (SSURGO) database (U.S. Department of Agriculture, 1995) was not available for all regions of interest in California. The STATSGO data were not suitable for use in this study in unprocessed form, so STATSGO data compiled by Schwarz and Alexander (1995) were used. These later data included weighted averaging of many of the soil characteristics contained in the database. The U.S. Department of Agriculture (1993) Soil Survey Manual provides more information on these soil characteristics:

- Soil slope (a percentage),
- Soil permeability (in in./hr),
- Soil organic matter content (a percentage, by weight),
- Soil erodibility factor (unitless),
- Soil hydrologic group (1 through 4; 1 has the highest infiltration rate),
- Soil drainage (1 through 7; 1 has the lowest drainage rate),
- Soil clay content (percentage of material with particle size less than 2 mm),
- Soil available water capacity (in in./in.),
- Soil thickness within the uppermost 79 inches below land surface (in in.), and
- Soil liquid limit (percentage of moisture by weight).

Using Logistic Regression Modeling to Predict the Probability of Debris Flows

All possible groupings and ordering of independent variables were evaluated to determine which combinations produced the most effective models. The models were built by sequentially adding variables to the analysis and evaluating McFadden's rho-squared, p-values calculated for each independent variable, and the sensitivity and specificity. Independent variables that did not increase model effectiveness were discarded.

Five logistic regression models were developed that incorporate a variety of independent variables (table 1). Percentage of high burn severity and 3-hr peak rainfall intensity were significant variables in all models. Soil organic matter content and soil clay content were significant variables in all models except model 5. Soil slope was a significant variable in all models except model 4. There are no significant differences in McFadden's rho-squared, sensitivity, or specificity among the five models, indicating that all five effectively predict the probability of debris flows. The most suitable model for the particular intended purpose can be selected from these five models on the basis of the availability of independent variables in the particular area of interest and field checking of probability maps.

Validation of Probability Models

The preferred method for validating the logistic regression models would be to compare them with an independent set of burned basins in southern California. Unfortunately, an independent set of burned basins is not yet available, although many areas of southern California burned during 2007 (Cannon and others, 2008) and could provide a useful data set if rainfall events occur in the near future. The number of basins evaluated in this study was insufficient to split into a calibration dataset and a validation dataset. To demonstrate model performance, the percentage of actual debris flows were plotted with the predicted probability of debris flows using a deciles-of-risk calculation (fig. 2). There were insufficient data points to plot all 10 decile points, so the data were grouped into 5 increments (0 to less than 0.1, 0.1 to less than

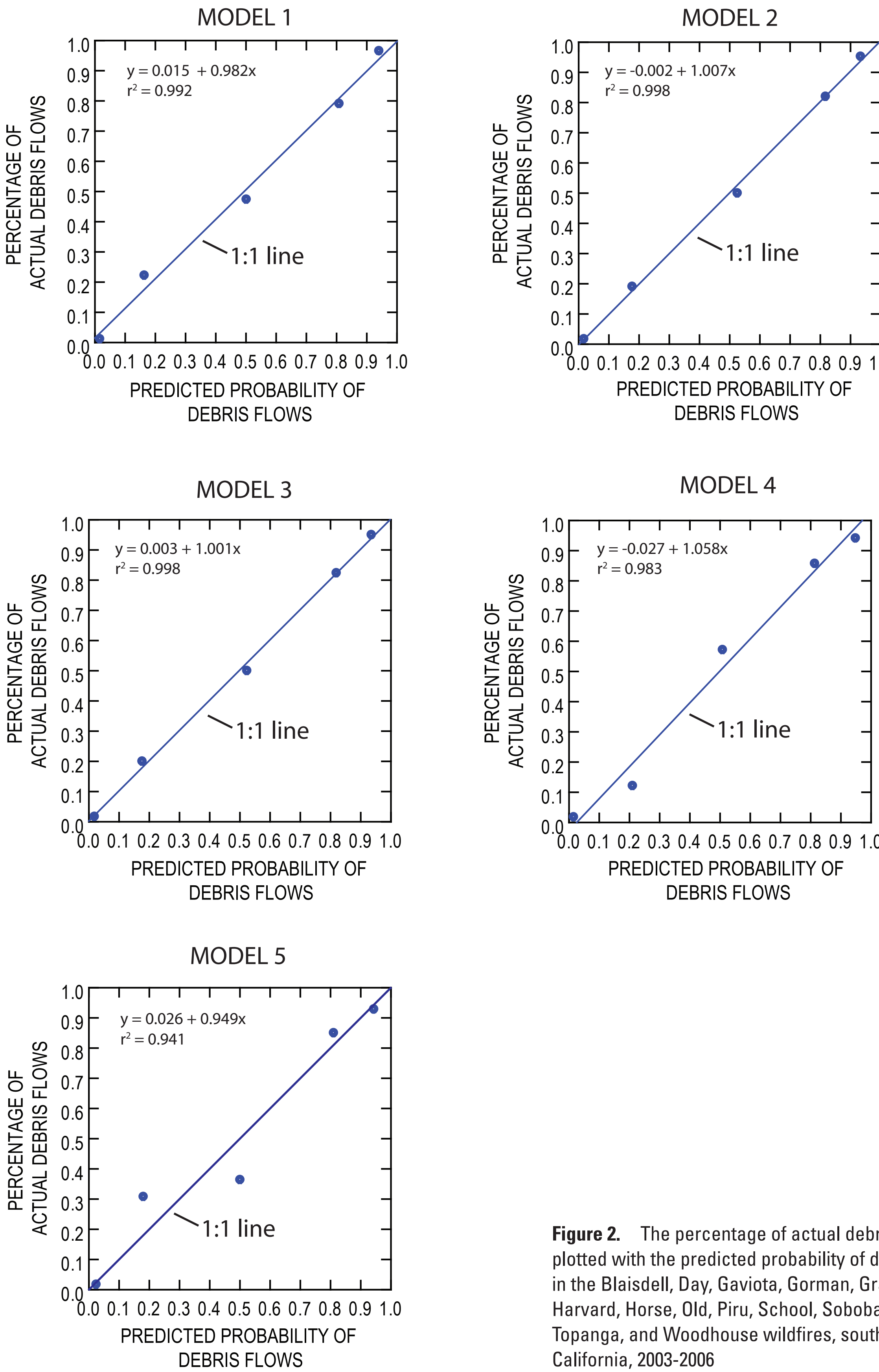
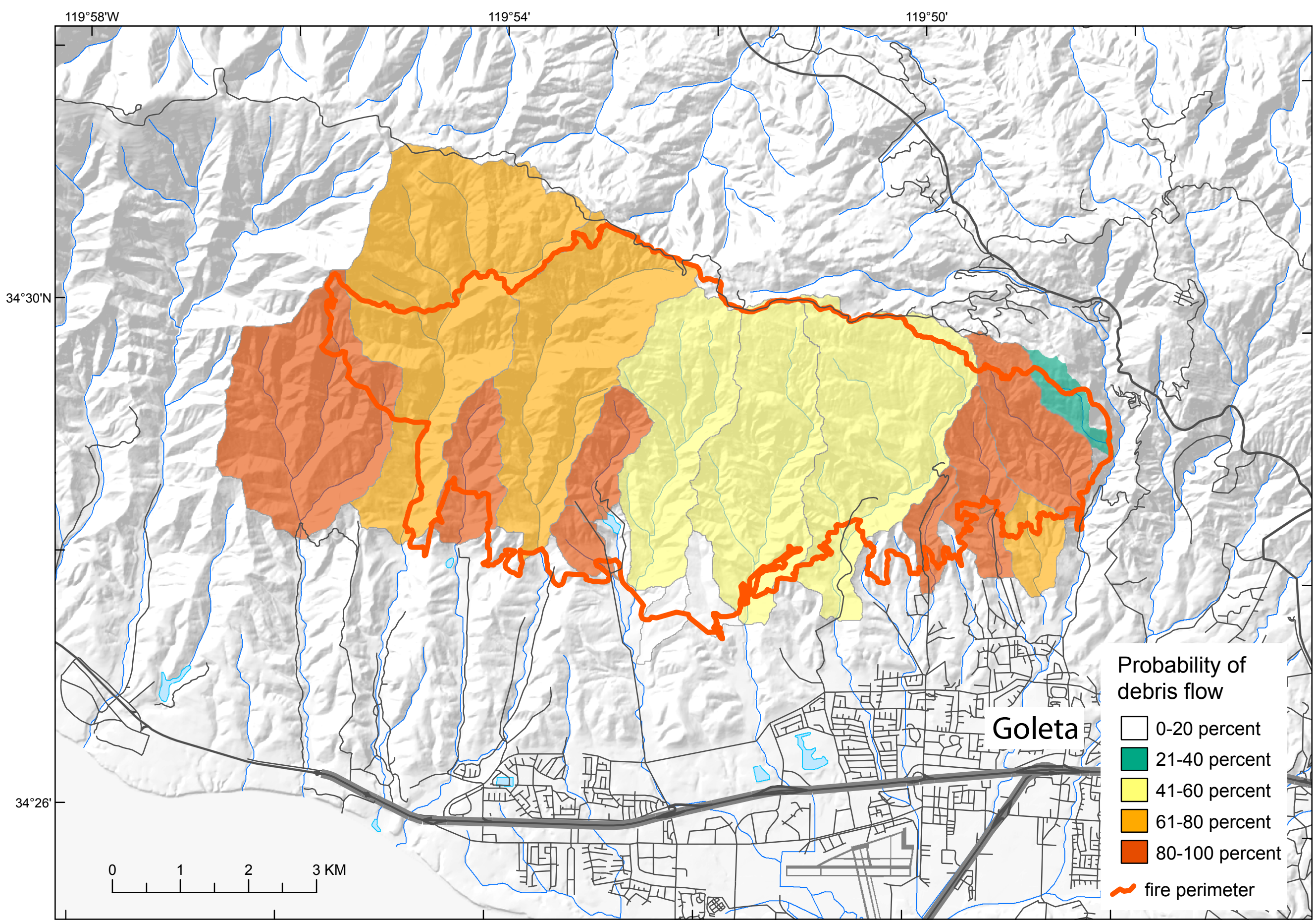


Figure 2. The percentage of actual debris flows plotted with the predicted probability of debris flows in the Blaisdell, Day, Gaviota, Gorman, Grand Prix, Harvard, Horse, Old, Piru, School, Soboba, Thurman, Topanga, and Woodhouse wildfires, southern California, 2003-2006

0.3, 0.3 to less than 0.7, 0.7 to less than 0.9, and 0.9 to 1), which gave each increment at least 20 data points. Figure 2 shows a good correlation between the percentage of actual debris flows and the predicted probability of debris flows. R-squared values of all five models were between 0.941 and 0.998, indicating that predicted probabilities of debris flows matched the actual percentage of debris flows very well. The specificity values were all greater than 0.9, indicating that the models correctly predicted areas where debris flows did not occur with greater than a 90 percent success rate (table 1). The sensitivity values were all greater than 0.74, indicating that the models correctly predicted where debris flows did occur about 75 percent of the time.

Applying Model 2 to the 2007 Gap Wildfire

To demonstrate the utility of the probability models developed in this study, a map showing the probability of debris flows occurring after the 2007 Gap Wildfire in southern California was generated using model 2 (table 1) for a 10-year recurrence, 3-hr duration rainstorm of 4.8 cm (1.9 in.) (fig. 3). The rainfall amount was determined by Hershfield (1961). Probabilities between 0.2 percent and 97.8 percent were calculated for the burned basins. On the map, the calculated probabilities are shown parsed into 20 percent probability classes, rather than as absolute values, to account for some of the uncertainty in the calculation.



Base from Environmental Systems Research Institute (ESRI) Data & Maps, original data extracted from the 2000 Census TIGER/Line® files 1:100,000, Universal Transverse Mercator projection, Zone 11 North American Datum of 1983

Figure 3. Probability of debris flows after a 10-year recurrence, 3-hour rainstorm of 4.8 centimeters (1.9 inches) at the Gap wildfire, southern California, within 3 years after the wildfire, using Model 2

Use and Limitations of the Models

The probability models developed in this study are designed to portray the potential, or predisposition, for post-wildfire debris flows in southern California drainage basins. The models do not identify the absolute certainty that a given basin will experience a debris flow, but rather the potential for (or likelihood of) a debris flow if a rainstorm of sufficient intensity occurs within 3 years after a wildfire. Probability is not the same as certainty; a high-probability basin may not experience a debris flow, and a low-probability basin may still experience a debris flow. Although the models calculate the probability of a debris flow, there is inherent uncertainty within these predictions.

The independent variables included in the models are considered to be possible first-order effects that can be rapidly evaluated immediately after a fire and are commonly available. Additional variables may also affect debris-flow production from recently burned basins in southern California. For example, an abundance of dry-ravel material in a specific channel may affect the volume of sediments transported by debris flows, and the frequently occurring fire-and-flood sequence that characterizes southern California basins may similarly limit material available for incorporation into debris flows (Spittler, 1995). However, data necessary to evaluate these effects are not currently (2008) available and are beyond the scope of this report.

References Cited

Brock, R.J., Cannon, S.H., Gartner, J.E., Santi, P.M., Higgins, J.D., and Bernard, D.R., 2007, An ordinary storm with an extraordinary response—Mapping the 2003 Old and Grand Prix Fire areas in southern California: Geological Society of America Abstracts with Programs, v. 39, no. 6, p. 180.

Cannon, S.H., 2000, Debris-flow response of southern California watersheds recently burned by wildfire, *in* Wieczorek, G.F., and Naeser, N.D., eds., Debris-flow hazards mitigation—Mechanics, prediction, and assessment, International Conference on Debris-Flow Hazards Mitigation, 2d., Taipei, Taiwan, 16–18 August 2000: Rotterdam, A.A. Balkema, Proceedings, p. 45–52.

Cannon, S.H., 2001, Debris-flow generation from recently burned watersheds: Environmental and Engineering Geoscience, v. 7, p. 321–341.

Cannon, S.H., 2005, Southern California—Wildfires and debris flows: U.S. Geological Survey Fact Sheet 2005–3106, accessed November 18, 2008, at URL *http://pubs.usgs.gov/fs/2005/3106/pdf/FS-3106.pdf*.

Cannon, S.H., Gartner, J.E., and Michael, J.A., 2007, Methods for the emergency assessment of debris flow hazards from basins burned by the fires of 2007, southern California: U.S. Geological Survey Open-File Report 2007-1384, 10 p., accessed December 16, 2008 at URL *http://pubs.usgs.gov/of/2007/1384/pdf/OF07-1384_508.pdf*.

Cannon, S.H., Gartner, J.E., Wilson, R.C., and Laber, J.L., 2008, Storm rainfall conditions for floods and debris flows from recently burned areas in southwestern Colorado and southern California: Geomorphology, v. 96, p. 250-269.

Cannon, S.H., Gartner, J.E., Rupert, M.G., Michael, J.A., Rea, A.H., and Parrett, Charles, in press, Predicting the probability and volume of post-wildfire debris flows in the Intermountain Western U.S.: The Geological Society of America Bulletin, approved for publication by the journal on October 12, 2008.

Chong, Jia-Rui, Renaud, Jean-Paul, Ailsworth, Erin, 2004, Flash floods wash away lives, dreams: Los Angeles Times, January 3, 2004, p. B1.

Helsel, D.R., and Hirsch, R.M., 2002, Studies in environmental science 49—Statistical methods in water resources: U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, chap. A3, 522 p.

Hershfield, D.M., 1961, Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years: Technical Paper No. 40, U.S. Department of Commerce, Weather Bureau, 61 p.

Horton, R.E., 1945, Erosional development of streams and their drainage basins — Hydrological approach to quantitative geomorphology: Geological Society of America, Bulletin, no. 56, p. 275–370.

Hosmer, D.W., and Lemeshow, Stanley, 2000, Applied logistic regression: New York, John Wiley, 375 p.

Iverson, R.M., 1997, The physics of debris flows: Reviews in Geophysics: v. 35, p. 245–296.

Key, C.H., and Benson, N.C., 2006, Landscape assessment—Remote sensing of severity, the normalized burn ratio, *in* Lutes, D.C., Keane, R.E., Caratti, J.F., Key, C.H., Benson, N.C. and Gangi, L.J., FIREMON—Fire effects monitoring and inventory system: United States Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, Utah, General Technical Report RMS–GTR–164–CD:LA1–LA51, p. LA-1 to LA-51 accessed on November 21, 2008, at URL *http://frames.nbii.gov/projects/firemon/FIREMON_LandscapeAssessment.pdf*.

Kleinbaum, D.G., 1994, Logistic regression, a self-learning text: New York, Springer-Verlag, 282 p.

Menard, Scott, 2002, Applied logistic regression analysis: Thousand Oaks, Calif., Sage Publications, 111 p.

Rupert, M.G., Cannon, S.H., and Gartner, J.E., 2003, Using logistic regression to predict the probability of debris flows occurring in areas recently burned by wildland fires: U.S. Geological Survey Open-File Report 03-500, poster, available only online at *http://water.usgs.gov/pubs/of/2003/ofr03500/*.

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, accessed September 10, 2007, at URL *http://water.usgs.gov/lookup/getspatial?ussoils*.

Spittler, T.E., 1995, Fire and debris flow potential of winter storms, *in* Keely, J.E., and Scott, T., Brushfires *in* California wildlands—Ecology and resource management: Fairfield, Wash., International Association of Wildland Fire, p. 113–120, accessed November 21, 2008, at URL *http://www.werc.usgs.gov/seki/pdfs/Splittler_Fire-and_Debris_Flow_Winter.pdf*.

Stevens, M.R., Bossong, C.R., Rupert, M.G., Ranalli, A.J., Cassidy, E.W., and Druliner, A.D., 2008, Post-wildfire hydrologic hazards in the wildland urban interface of Colorado and the Western United States: U.S. Geological Survey Fact Sheet 2007–3036, 6 p., accessed November 17, 2008, at URL *http://pubs.usgs.gov/fs/2007/3036/pdf/FS07-3036_508.pdf*.

SYSTAT Software, Inc., 2004, SYSTAT 11, Statistics II—Software documentation: Richmond, Calif., SYSTAT Software, Inc., 657 p.

U.S. Department of Agriculture, 1991, State Soil Geographic (STATSGO) data base: U.S. Department of Agriculture, Soil Conservation Service Miscellaneous Publication 1492, 88 p.

U.S. Department of Agriculture, 1993, Soil survey manual: U.S. Department of Agriculture Handbook 18, 437 p.

U.S. Department of Agriculture, 1995, Soil survey geographic (SSURGO) data base: U.S. Department of Agriculture, Natural Resources Conservation Service Miscellaneous Publication 1527, 31 p.

U.S. Geological Survey, 1999, National elevation dataset: U.S. Geological Survey Fact Sheet 148-99, accessed December 16, 2008, at URL *http://egsc.usgs.gov/isb/pubs/factsheets/fs14899.html*.

Rupert and others—**Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, California, 2003–2006**—OFR 2008–1370