



Updated Logistic Regression Equations for the Calculation of Post-Fire Debris-Flow Likelihood in the Western United States

By Dennis M. Staley, Jacquelyn A. Negri, Jason W. Kean, Jayme M. Laber, Anne C. Tillery, and Ann M. Youberg

Open-File Report 2016–1106

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1-888-ASK-USGS (1-888-275-8747).

For an overview of USGS information products, including maps, imagery, and publications, visit <http://store.usgs.gov/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Staley, D.M., Negri, J.A., Kean, J.W., Laber, J.M., Tillery, A.C., and Youberg, A.M., 2016, Updated logistic regression equations for the calculation of post-fire debris-flow likelihood in the Western United States: U.S. Geological Survey Open-File Report 2016-1106, 13 p., <http://dx.doi.org/10.3133/ofr20161106>.

Contents

Abstract 1
 Introduction..... 1
 Methods..... 4
 Results..... 8
 Conclusions 10
 Acknowledgments 10
 References Cited 10
 Appendix 1..... 13

Figures

1. Recently burned areas that compose the training dataset and test dataset for the Western United States and southern California 5
 2. Receiver operating characteristic (ROC) analysis methods and metrics demonstrating the 2x2 contingency table identifying the four possible outcomes of a binary classifier model and the threat score measurement of classifier performance used in ROC analysis..... 8

Tables

1. Logistic regression variables and coefficient values for the link functions of models published previously and in this study for predicting the statistical likelihood of post-fire debris-flow generation in the Western United States 3
 2. Fires, number of records, and number of debris flows for data included in training and testing datasets 6
 3. Regression evaluation and classification evaluation metrics for five analyzed logistic regression models..... 9

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
meter (m)	1.094	yard (yd)
Area		
square kilometer (km ²)	247.1	acre
square kilometer (km ²)	0.3861	square mile (mi ²)

Datum

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88). Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83). Altitude, as used in this report, refers to distance above the vertical datum, in meters.

Abbreviations

AIC _c	Akaike Information Criterion
BARC	Burned Area Reflectance Classification
dNBR	difference normalized burn ratio
IMW	Intermountain West
R ²	coefficient of determination
ROC	receiver operating characteristics
SoCal	southern California
TS	threat score
USGS	U.S. Geological Survey

Updated Logistic Regression Equations for the Calculation of Post-Fire Debris-Flow Likelihood in the Western United States

By Dennis M. Staley,¹ Jacquelyn A. Negri,¹ Jason W. Kean,¹ Jayme M. Laber,² Anne C. Tillery,¹ and Ann M. Youberg³

Abstract

Wildfire can significantly alter the hydrologic response of a watershed to the extent that even modest rainstorms can generate dangerous flash floods and debris flows. To reduce public exposure to hazard, the U.S. Geological Survey produces post-fire debris-flow hazard assessments for select fires in the Western United States. We use publicly available geospatial data describing basin morphology, burn severity, soil properties, and rainfall characteristics to estimate the statistical likelihood that debris flows will occur in response to a storm of a given rainfall intensity. Using an empirical database and refined geospatial analysis methods, we defined new equations for the prediction of debris-flow likelihood using logistic regression methods. We showed that the new logistic regression model outperformed previous models used to predict debris-flow likelihood.

Introduction

Since 1970, the frequency of large wildfires, length of fire season, and duration of individual wildfires have steadily increased in the Western United States as a result of a combination of human activities and evolving land-use patterns, weather, and climate (Westerling and others, 2006). Increases in the susceptibility to debris flows are a secondary effect of wildfire in recently burned steeplands, a hazard which may persist for several years following fire containment (Cannon and DeGraff, 2009; Cannon and others, 2010; DeGraff and others, 2015). Risk associated with debris-flow hazards increases as populations continue to expand into foothill and mountainous areas susceptible to wildfire. In addition, a greater incidence of fire activity in mountainous areas with relatively infrequent fire recurrence may increase the potential of debris flows in environments or communities where debris-flow hazard has been historically absent (Cannon and DeGraff, 2009). This geographic expansion of areas exposed to post-fire debris-flow hazard has motivated efforts to reduce exposure of people, infrastructure, and important natural, cultural, and economic resources to these hazards. Reducing exposure to such hazard is most commonly done through the development and

¹ U.S. Geological Survey

² National Weather Service

³ Arizona Geological Survey

improvement of methods used to predict the occurrence of debris flows for hazard assessment and early warning (Cannon and others, 2008, 2010, 2011; Staley and others, 2013a, c).

Logistic regression models are frequently used to predict the likelihood of post-fire debris flows (for example, Cannon and others, 2009, 2010; Staley and others, 2013a, c; Staley, 2014; U.S. Geological Survey, 2016). This approach to characterizing debris-flow hazard typically utilizes empirical data associated with debris-flow occurrence (or non-occurrence), rainstorm characteristics, and geospatial data describing basin morphology, burn severity, and soil properties. Cannon and others (2010) described a method used to calculate the statistical likelihood of post-fire debris-flow occurrence in the intermountain Western United States using a logistic regression model. Their technique incorporated data of past debris-flow occurrence combined with rainfall intensity data and geospatial data characterizing basin morphometry, burn severity, and soil properties to calculate the likelihood that a post-fire debris flow will occur given a rainfall intensity associated with a rainstorm of a known recurrence interval. Staley and others (2013a) used the methods and data from Rupert and others (2008) and Cannon and others (2010) and incorporated additional data from southern California burn areas where fires occurred between 2006 and 2010 to develop the most current model of statistical likelihood for southern California (Staley and others, 2013a). Both models are currently being applied operationally by the U.S. Geological Survey (2016).

Event occurrence is modeled using the logistic curve to define the statistical likelihood of a binary response (that is, debris-flow generation) as

$$P = \frac{e^x}{1 + e^x} \quad (1)$$

where P is a number ranging from 0 to 1 and represents the statistical likelihood of debris-flow occurrence (where values approaching 1 indicate an increasing likelihood) and x is the link function. The link function is defined according to the equation

$$x = \beta + C_1X_1 + C_2X_2 + \dots + C_nX_n \quad (2)$$

where β and C_1, C_2, \dots, C_n are empirically derived parameters and X_1, X_2, \dots, X_n represent independent variables that influence the occurrence of the event.

The current logistic regression equations are not suitable for defining a realistic relation between rainfall intensity and debris-flow likelihood. A realistic prediction of debris-flow generation requires that post-fire debris-flow likelihood should be close to zero in the absence of rainfall. In the current predictive models (Cannon and others, 2010; U.S. Geological Survey, 2016), rainfall variables influence the link function independently of the variables associated with basin morphology, burn severity, or soil properties. As such, values of P greater than zero occur in the existing likelihood equations (Cannon and others, 2010; U.S. Geological Survey, 2016), even when it is not raining. We accomplished improved realism in model predictions through the multiplicative combination of rainfall accumulation with variables related to basin morphology, fire severity, and soil properties, such that the link function for the new logistic model follows the equation

$$x = \beta + C_1X_1 + C_2X_2 + C_3X_3 \quad (3)$$

where $\beta, C_1, C_2,$ and C_3 are empirically defined parameters. This link function can then be used to calculate the statistical likelihood of debris-flow occurrence using equation 1.

The logistic regression approach to predicting post-fire debris-flow likelihood is advantageous as it is computationally simple, utilizes publicly available geospatial data, and can be designed to provide a statistical probability of the occurrence of debris flows for storms of different magnitudes in geospatial format at the scale of a stream segment or drainage basin (for example, Tillery and others, 2012; Verdin and others, 2012; U.S. Geological Survey, 2016). Currently, the U.S. Geological Survey (USGS) uses two models based on logistic regression to predict the likelihood of debris-flow occurrence in the Western United States (U.S. Geological Survey, 2016): the Intermountain West (IMW) model and the southern California (SoCal) model. The current model equations used to predict post-fire debris-flow likelihood in the Western United States can be found in table 1.

Table 1. Logistic regression variables and coefficient values for the link functions of models published previously (Cannon and others, 2010; Staley and others, 2013a) and in this study (Model M1) for predicting the statistical likelihood of post-fire debris-flow generation in the Western United States. The Model M1 intercept (β) and coefficient (C_1 , C_2 , and C_3) values are listed for model predictions based upon rainfall durations of 15, 30, and 60 minutes (min).

[\geq , greater than or equal to; %, percent; dNBR, difference normalized burn ratio; KF, Soil KF-Factor; --, not applicable]

Source	Model		
	Intermountain Western United States (A)	Southern California	M1 (15, 30, 60 min)
β	Cannon and others (2010) -0.7	Staley and others (2013a) -5.22	This study -3.63, -3.61, -3.21
C_1	0.03	0.003	0.41, 0.26, 0.17
X_1	Percentage of basin area with gradients $\geq 30\%$	Relief, in meters	Proportion of upslope area with moderate to high dNBR values and gradients $\geq 23^\circ$ \times Rainfall accumulation
C_2	-1.6	0.008	0.67, 0.39, 0.20
X_2	Ruggedness	Percentage of basin area burned at high or moderate severity and gradients $\geq 50\%$	(dNBR \div 1,000) \times Rainfall accumulation
C_3	0.06	0.024	0.70, 0.50, 0.220
X_3	Percentage of basin area burned at high or moderate severity	Average gradient of burned terrain, in percent	Soil KF-Factor \times Rainfall accumulation
C_4	0.2	-0.007	--
X_4	Soil clay content, in percent	Soil clay content, in percent	--
C_5	-0.4	0.105	--
X_5	Liquid limit	Peak 30-minute storm intensity, in millimeters per hour	--
C_6	0.07	--	--
X_6	Average storm intensity, in millimeters per hour	--	--

In this study, we present a new logistic regression model for the prediction of post-fire debris-flow likelihood (M1 in table 1) that objectively provides better predictions of post-fire debris-flow occurrence throughout the Western United States. In addition, the presented method is demonstrated to perform better than existing models in areas that were not used for original model calibration. Model calibration data, hereafter referred to as the “training dataset,” were collected in southern California (fig. 1). Model validation data, hereafter referred to as the “test dataset,” were collected in other areas of the Western United States (fig. 1). We compare the predictions made by the new model to those of the two existing models. Our results demonstrate that the M1 model equation provides improved predictions of post-fire debris-flow likelihood in the Western United States.

Methods

We relied on the established methods of logistic regression (equations 1 and 2) and receiver operating characteristics (ROC) analysis (Swets, 1988; Fawcett, 2006) to define new equations for predicting post-fire debris-flow likelihood. Logistic regression models were based upon empirical data collected within the first two years of wildfire in recently burned areas in the Western United States (fig. 1, table 2). In total, the database used for this study consisted of 1,550 records and included information pertaining to location, hydrologic response (debris flow or no debris flow), rainfall rates, and the morphological properties of the contributing area. Drainage areas for locations included in the database ranged from 0.2 to 8 square kilometers. Rainfall and response data, along with geographic information systems (GIS) metrics used in the modeling, may be found in appendix 1.



Figure 1. Recently burned areas that compose the training dataset (blue dots) and test dataset (red dots) for (A) the Western United States and (B) southern California.

Table 2. Fires, number of records, and number of debris flows for data included in training and testing datasets.

Fire name	Abbreviation	State	Year	Number of records	Number of debris flows
Training dataset					
Buckweed	bck	California	2007	16	0
Blaisdell	bla	California	2005	10	0
Canyon	can	California	2007	14	0
Day	day	California	2006	8	0
Freeway	fwy	California	2008	8	0
Gap	gap	California	2008	10	0
Grand Prix—Old	gpo	California	2003	78	60
Harris	har	California	2007	10	0
Horse	hrs	California	2006	9	0
Harvard	hrv	California	2005	28	6
Jesusita	jes	California	2009	6	0
Poomacha	poo	California	2007	21	19
Santiago	san	California	2007	12	5
Sayre	say	California	2008	14	2
School	sch	California	2005	12	0
Sesnon	ses	California	2008	4	0
Station	stn	California	2009	600	108
Topanga	top	California	2005	33	0
Witch	wit	California	2007	46	1
Test dataset					
Bear	br	Montana	2000	14	11
Cerro Grande	cg	New Mexico	2000	11	5
Coal Seam	cs	Colorado	2002	253	17
Fridley Peak	fri	Montana	2001	11	5
Gladiator	gld	Arizona	2012	35	3
Horseshoe 2*	h2f	Arizona	2001	30	4
Little Bear	lb	New Mexico	2012	47	30
Monument	mmt	Arizona	2011	19	6
Mollie	mol	Utah	2001	4	4
Missionary Ridge	mr	Colorado	2002	16	11
Purdy	pur	Montana	2001	9	0
Schultz*	scz	Arizona	2010	105	26
Sula	sul	Montana	2000	6	2
Waldo Canyon	wal	Colorado	2012	31	7
Wallow	wlw	Arizona	2011	20	2

*Differenced normalized burn ratio (dNBR) data not available as a result of sensor malfunction.

Rainfall data were collected at nearby rain gages (maximum distance of 4 kilometers) from a variety of sources, including the USGS, the National Oceanographic and Atmospheric Administration, and other State, county, and local agencies. Rainfall accumulations and peak storm intensities were calculated using a backwards differencing approach (Kean and others, 2011). For this study, individual storms were defined by intervals of at least 8 hours without rainfall. Hydrologic response information (debris flow or no debris flow) was collected by USGS personnel and local collaborators.

The properties of the contributing area for each record were derived from publicly available geospatial data sources. Although higher-resolution data sources may be available for specific locations, we chose to use the following data sources because they are available nationwide, allowing for consistent calculation of metrics for any burn area in the United States. Topographic data used to calculate morphologic variables were derived from 10-meter digital elevation models (U.S. Geological Survey, 2015), and soil property data were extracted from the USGS State Soil Geographic (STATSGO) database (Schwartz and Alexander, 1995). Burn severity information was provided by local Burned Area Emergency Response (BAER) teams, who field-validated Burned Area Reflectance Classification (BARC) imagery derived from the differenced normalized burn ratio (dNBR) (Key and Benson, 2006).

The training dataset consisted of 939 records from southern California, of which 201 were from debris-flow producing rainstorms. The test dataset consisted of 611 total records (133 debris-flow events) from other areas in the Western United States. While other methods of model calibration and validation were evaluated (such as randomly split datasets and bootstrapping), we divided the dataset in this manner for three reasons. First, the training and test datasets had similar ranges in variable values for the proposed model. We consider the data included in the training dataset to be representative of the range of variable values found in the test dataset. Second, this method of geographic division of the data provided the greatest degree of model accuracy on both training and test datasets, as measured by regression evaluation statistics and model classification evaluation methods. Finally, the data included in the training dataset were collected as a part of a broader USGS Landslide Hazards Program monitoring effort in southern California. We consider these data to be of the highest quality with regards to rain gage accuracy, proper identification of debris-flow response, and location of response. Data included in the test dataset were also of high quality and were checked for accuracy, but different data sources (such as different types of rain gages and different observers of debris-flow response) may have resulted in slight inconsistencies in rainfall and response characterization.

The training dataset was used to develop the logistic regression model. Hereafter, the term “logistic model” refers to the set of equations used to predict debris-flow likelihood at durations of 15, 30, and 60 minutes. The logistic model was composed of multiple equations with the same variables related to basin morphology, fire severity, and soil properties and differing parameter values (β , C_1 , C_2 , and C_3) for rainfall accumulations (in millimeters) measured over the analyzed duration.

Model fit was evaluated against the training dataset using regression evaluation metrics that include the Adjusted R^2 , the Tjur R^2 , and the corrected Akaike Information Criterion (AIC_c) (Akaike, 1974; Tjur, 2009; Negri, 2016). These metrics were selected for inclusion as they have been identified as being useful for the evaluation of predictive models using logistic regression (Negri, 2016). For the two R^2 metrics, higher values represent better model fit of the training data. Lower values of AIC_c represent better model fit of the training data, although AIC_c values cannot be directly compared when the number of model variables or records is different. Data characterizing the fit of the original models (IMW and SoCal) were not reported in the model publications, so they are not reported here.

		OBSERVED CLASS	
		Debris flow	No debris flow
MODELED CLASS	Above threshold	True positive (TP) <i>Above threshold, debris flow observed</i>	False positive (FP) <i>Above threshold, no debris flow observed</i>
	Below threshold	False negative (FN) <i>Below threshold, debris flow observed</i>	True negative (TN) <i>Below threshold, no debris flow observed</i>

$$\text{Threat score} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$$

Figure 2. Receiver operating characteristic (ROC) analysis methods and metrics demonstrating the 2×2 contingency table identifying the four possible outcomes of a binary classifier model and the threat score measurement of classifier performance used in ROC analysis.

Model performance was then evaluated against both the training dataset and the test dataset, as well against all data combined, using the threat score (TS) metric (table 3). The threat score is a classifier evaluation metric from receiver operating characteristics (ROC) analysis (Swets, 1988; Fawcett, 2006) commonly used in post-fire debris-flow studies (Staley and others, 2013b, 2015; Youberg, 2014). The threat score is based upon the distribution of model predictions and observed outcomes in a confusion matrix (fig. 2), where a perfect model would have a TS equal to 1, and each incorrect prediction (false positive or false negative in figure 2) would reduce the model TS value.

Results

The M1 equations (M1_15, M1_30, and M1_60) represent a logistic regression model capable of predicting debris-flow likelihood for storm intensities measured over durations of 15, 30, and 60 minutes (Model M1, table 1). Statistical and ROC performance measures of model performance for the training and test datasets are reported in table 3.

Five equations were considered: three based on the M1 model and one each based on the IMW and SoCal models. In terms of the TS metric, the M1 15-minute-duration model (M1_15) outperformed the current IMW and SoCal models for all records in the database and for the test dataset (table 3). The SoCal model outperformed M1_15 in TS for the training dataset, though only by a very small margin (TS of 0.43 for SoCal compared to 0.42 for M1_15). In terms of the Tjur R^2 metric, M1_15 outperformed the other models for all three datasets. For the new equations (M1_15, M1_30, and M1_60), analysis at the 15-minute duration proved to provide the best predictions of post-fire debris-flow occurrence.

Table 3. Regression evaluation and classification evaluation metrics for five analyzed logistic regression models.

[Values in bold indicate the model with the highest metric value (TS, AIC_c, Adjusted R², Tjur R²) for each dataset. The regression evaluation metrics for the Intermountain West (IMW) model (Cannon and others, 2010) and the southern California (SoCal) model (Staley and others, 2013) were not recorded in their original publications, so they are not reported here. Equations from the model presented in this report (M1) are based on durations of 15 minutes, 30 minutes, and 60 minutes (M1_15, M1_30, and M1_60, respectively). AIC_c, Akaike Information Criterion; R², coefficient of determination; --, not applicable]

Dataset	Model	Number of records	Number of debris flows	True negative (TN)	False negative (FN)	False positive (FP)	True positive (TP)	Threat score (TS)	AIC _c	Adjusted R ²	Tjur R ²
All records	M1_15	1,243	316	766	125	161	191	0.40	--	--	0.35
	M1_30	1,243	316	836	215	91	101	0.38	--	--	0.34
	M1_60	1,243	316	888	273	39	43	0.34	--	--	0.30
	SoCal	1,336	328	820	135	188	193	0.37	--	--	0.30
	IMW	1,390	316	1,069	305	5	11	0.03	--	--	0.03
Training	M1_15	815	201	584	105	30	96	0.42	643.0	0.42	0.34
	M1_30	815	201	609	171	5	30	0.42	632.5	0.43	0.35
	M1_60	815	201	614	194	0	7	0.39	642.0	0.41	0.33
	SoCal	805	201	551	91	53	110	0.43	--	--	0.33
	IMW	819	201	618	201	0	0	0.00	--	--	0.00
Test	M1_15	428	115	182	20	131	95	0.39	--	--	0.33
	M1_30	428	115	227	44	86	71	0.35	--	--	0.32
	M1_60	428	115	274	79	39	36	0.26	--	--	0.27
	SoCal	531	127	269	44	135	83	0.32	--	--	0.25
	IMW	571	115	451	104	5	11	0.09	--	--	0.10

Conclusions

This report describes a new, fully predictive framework for assessing post-fire debris-flow hazards using publicly available geospatial data. Our approach can be used to predict the likelihood of debris-flow occurrence for recently burned areas where there are no preexisting historical data concerning debris-flow generation. Specifically, the M1 15-minute-duration (M1_15) logistic model is recommended for use in the Western United States (fig. 1A), as it objectively produced better predictions of debris-flow occurrence in both the training dataset and the test dataset when compared to the current logistic models and M1 at durations of 30 and 60 minutes. The methods presented in this study are applicable for locations situated in recently burned areas for a period of one to two years following a wildfire. Additional research, including long-term post-fire monitoring, would enable us to better constrain the relation between recovery of the vegetation and soil systems and the reduction of debris-flow susceptibility, as well as ultimately the increasing rainfall intensity required to generate debris flows in older burn areas. The models presented here should not be applied to unburned areas or areas where post-fire debris flows are commonly generated by shallow landslides.

Acknowledgments

We are grateful for assistance from Anne-Marie Matherne, Joseph Gartner, Kevin Schmidt, Anne-Marie Matherne, Maiana Hanshaw, Robert Leeper, Octaviano Lucero, Pete Wohlgemuth, and Terri Hogue.

References Cited

- Akaike, H., 1974, A new look at the statistical model identification: Institute of Electrical and Electronics Engineers (IEEE) Transactions on Automatic Control, v. 19, no. 6, p. 716–723.
- Cannon, S.H., Boldt, E.M., Laber, J.L., Kean, J.W., and Staley, D.M., 2011, Rainfall intensity–duration thresholds for postfire debris-flow emergency-response planning: Natural Hazards, v. 59, no. 1, p. 209–236.
- Cannon, S.H., and DeGraff, J.V. 2009, The increasing wildfire and post-fire debris-flow threat in western USA, and implications for consequences of climate change, *in* Sassa, K., and Canuti, P., eds., Landslides—Disaster risk reduction: Berlin, Heidelberg, Springer, p. 177–190.
- Cannon, S.H., Gartner, J.E., Rupert, M.G., Michael, J.A., Rea, A.H., and Parrett, C., 2010, Predicting the probability and volume of postwildfire debris flows in the intermountain Western United States: Geological Society of America Bulletin, v. 122, no. 1–2, p. 127–144.
- Cannon, S.H., Gartner, J.E., Rupert, M.G., Michael, J.A., Staley, D.M., and Worstell, B.B., 2009, Emergency assessment of postfire debris-flow hazards for the 2009 Station Fire, San Gabriel Mountains, southern California: U.S. Geological Survey Open-File Report 2009–1227, 20 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2009/1227/>.
- Cannon, S.H., Gartner, J.E., Wilson, R.C., Bowers, J.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, no. 3–4, p. 250–269.
- DeGraff, J.V., Cannon, S.H., and Gartner, J.E., 2015, Timing of susceptibility to post-fire debris flows in the western USA: Environmental & Engineering Geoscience, v. 21, no. 4, p. 277–292.
- Fawcett, T., 2006, An introduction to ROC analysis: Pattern Recognition Letters, v. 27, p. 861–874.

- Kean, J.W., Staley, D.M., and Cannon, S.H., 2011, In situ measurements of post-fire debris flows in southern California—Comparisons of the timing and magnitude of 24 debris-flow events with rainfall and soil moisture conditions: *Journal of Geophysical Research*, v. 116, no. F4, article F04019.
- Key, C.H., and Benson, N.C., 2006, Landscape Assessment (LA) sampling and analysis methods: U.S. Department of Agriculture–Forest Service General Technical Report RMRS–GTR–164, 51 p.
- Negri, J.A., 2016, Evaluation and validation of multiple predictive models applied to post-wildfire debris-flow hazard prediction: Golden, Colo., Colorado School of Mines, M.S. thesis, 94 p.
- 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, 16 p., accessed May 4, 2016, at <https://pubs.er.usgs.gov/publication/ofr20081370>.
- Schwartz, G.E., and Alexander, R.B., 1995, Soils data for the conterminous United States derived from the NRCS State Soil Geographic (STATSGO) Database: U.S. Geological Survey Open-File Report 95–449, accessed May 4, 2016, at <http://water.usgs.gov/lookup/getspatial?/ussoils>.
- Staley, D.M., 2014, Emergency assessment of post-fire debris-flow hazards for the 2013 Springs Fire, Ventura County, California: U.S. Geological Survey Open-File Report 2014–1001, 10 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2014/1001/>.
- Staley, D.M., Gartner, J.E., and Kean, J.W., 2015, Objective definition of rainfall intensity-duration thresholds for post-fire flash floods and debris flows in the area burned by the Waldo Canyon fire, Colorado, USA, *in* Lollino, Giorgio; Giordan, Daniele; Crosta, G.B.; Corominas, Jordi; Azzam, Rafiq; Wasowski, Janusz; Sciarra, Nicola, eds., *Landslide processes. Engineering geology for society and territory*, v. 2: Cham, Switzerland, Springer International Publishing, p. 621–624.
- Staley, D.M., Gartner, J.E., Smoczyk, G.M., and Reeves, R.R., 2013a, Emergency assessment of post-fire debris-flow hazards for the 2013 Mountain Fire, southern California: U.S. Geological Survey Open-File Report 2013–1249, 13 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2013/1249/>.
- Staley, D.M., Kean, J.W., Cannon, S.H., Schmidt, K.M., and Laber, J.L., 2013b, Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California: *Landslides*, v. 10, no. 5, p. 547–562.
- Staley, D.M., Smoczyk, G.M., and Reeves, R.R., 2013c, Emergency assessment of post-fire debris-flow hazards for the 2013 Powerhouse Fire, southern California: U.S. Geological Survey Open-File Report 2013–1248, 13 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2013/1248/>.
- Swets, J.A., 1988, Measuring the accuracy of diagnostic systems: *Science*, v. 240, no. 4857, p. 1285–1293.
- Tillery, A.C., Matherne, A.M., and Verdin, K.L., 2012, Estimated probability of postwildfire debris flows in the 2012 Whitewater–Baldy Fire burn area, southwestern New Mexico: U.S. Geological Survey Open-File Report 2012–1188, 11 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2012/1188/>.
- Tjur, T., 2009, Coefficients of determination in logistic regression models—A new proposal. The coefficient of discrimination: *The American Statistician*, v. 63, no. 4, p. 366–372.

- U.S. Geological Survey, 2015, The National Map—3D Elevation Program: U.S. Geological Survey Web page, accessed May 4, 2016, at http://nationalmap.gov/3dep_prodserv.html.
- U.S. Geological Survey, 2016, Emergency assessment of post-fire debris-flow hazards: U.S. Geological Survey Landslide Hazards Program Web page, accessed May 4, 2016, at http://landslides.usgs.gov/hazards/postfire_debrisflow/.
- Verdin, K.L., Dupree, J.A., and Elliot, J.G., 2012, Probability and volume of potential postwildfire debris flows in the 2012 Waldo Canyon burn area near Colorado Springs, Colorado: U.S. Geological Survey Open-File Report 2012–1158, 8 p., accessed May 4, 2016, at <http://pubs.usgs.gov/of/2012/1158/>.
- Westerling, A.L., Hidalgo, H.G., Cayan, D.R., and Swetnam, T.W., 2006, Warming and earlier spring increase western U.S. forest wildfire activity: *Science*, v. 313, no. 5789, p. 940–943.
- Youberg, A.M., 2014, Modern and ancient debris flows in Arizona: Tucson, Ariz., University of Arizona, Ph.D. dissertation, 235 p.

Appendix 1. Data Supporting Logistic Regression Model Calibration and Testing

Training and testing data used to calibrate and test the logistic regression model M1. Dataset includes information pertaining to event location, date, storm characteristics, basin morphology, fire severity, and soil properties.

The Excel file can be accessed from the report Web page (<http://dx.doi.org/10.3133/ofr20161106>). The first tab, “DataDescription,” defines abbreviations and explains column headings. Data appear on the second tab, “Appendix1_ModelData.”

ISSN 2331-1258 (online)
<http://dx.doi.org/10.3133/ofr20161106>