Incorporating Shear Stiffness into Post-Fire Debris Flow Triggering Models

ROBB MOSS (✉ rmoss@calpoly.edu)

California Polytechnic State University <https://orcid.org/0000-0003-1104-1476>

Noah Lyman

Corps: US Army Corps of Engineers

Research Article

Keywords: debris flow, shear wave velocity, predictive modeling

Posted Date: December 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1178662/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Incorporating Shear Stiffness into Post-Fire Debris Flow Triggering Models

R.E.S. Moss¹ and N. Lyman²

¹Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, CA, 93407.

²U.S. Army Corps of Engineers, Seattle District, Seattle, WA, 98134

Corresponding author: Robb Moss (rmoss@calpoly.edu)

Key Points:

- Shear stiffness of sediment controls debris flow triggering likelihood.
- At map scale the thirty-meter shear wave velocity is a proxy for shear stiffness.
- Debris flow triggering models can incorporate thirty-meter shear wave velocity for improved predictions.

Key Words: debris flow, shear wave velocity, predictive modeling

Abstract

Current post-fire debris flow triggering models consider predictor variables accounting for; rainfall intensity, rainfall accumulation, area burned, burned intensity, geology, slope, and others. These models represent the physical process of debris flow initiation and subsequent shear failure by quantifying near-surface soil characteristics. By including shear wave velocity as a proxy for sediment shear stiffness, models can better inform the likelihood of particle dislocation, contractive or dilative volume changes, and downslope displacement that results from debris flows. This broadly available variable common to other hazard predictions, such as liquefaction analysis, provides good coverage in the watersheds of interest for debris flow predictions. A logistic regression is used to compare the new variable against currently used variables for predictive post-fire debris flow triggering models. We find that the new variable produces improved performance in prediction of triggering while capturing the physics of sediment failing in a shearing and flow-type response. Additional suggestions are presented for utilizing statistical cross-validation methods to advance prediction performance, and the utility of different variables for quick assessment of likelihood during eminent high intensity rainfall events.

1 Introduction

As the Western United States continues to be exposed to a higher frequency of wildfires with increasing severity, development has also moved into steeper and more mountainous terrain [Westerling et al., 2006]. Add high-intensity rainfalls to this situation and the outcome is an increased recurrence of hazards such as debris flows [Cannon and DeGraff, 2009]. When critical infrastructure, businesses, and homes intersect with debris flows the impacts tend to result in high economic and social impact [Dowling and Santi, 2014; Kean et al., 2019].

Referred to at times as the “fire-flood cycle” [Kotok, 1935], runoff generated debris flows often follow wildfires due to alterations in the surface soil and hydrologic conditions in a watershed. Some modifications to these systems include water infiltration reduction [McGuire et al., 2018], alterations to local hydrographs [Rowe et al., 1949], and the reduction of potential soil stabilizing vegetation [De Graff, 2018]. These changes tend to increase the number of debris flows in many areas of increased development in Southern California and beyond.

To address this increased risk, efforts to predict and mitigate post-fire debris has been worked on by a cooperative program run by United States Geologic Survey (USGS) and National Oceanic and Atmospheric Agency (NOAA) Debris-Flow Warning System [USGS, 2005], and also by researchers investigating such topics as geomorphic mechanisms of debris flows in post-fire situations (eg. [McGuire et al., 2018]). Such work typically relies on a combination of meteorological forecasting and physical characteristics of post-fire watersheds in statistical models to forecast the probability of occurrence over differing rainfall scenarios.

Improvement upon forecasting both increases our understanding of the fundamental processes that govern debris flows while also permitting risk to be better managed. Being that debris flows are at times considered stochastic systems of varying terrains, lithologies, and hydrologic regimes, a further understanding of these fundamental variables could also reduce the data

processing, consolidate data availability through space and time, and increase adaptation of risk prediction models into new areas.

In this study we analyze logistic regression based statistical models in the assessment of debris flow probability by examining the input variables proposed by the USGS and others. A Southern California database of known post-fire rainfall induced debris flow observations are used to assess the performance of using 30m shear wave velocity (V_{s30}), in place of Soil-KF Factor, a measure of soil erodibility. Previous studies on initiation of post-fire debris flows suggest that surface variables relating to fundamental soil characteristics can improve upon current prediction methods while also providing a path forward in assessing other aspects of hazard such as runout length [Gartner et al., 2014].

2 Post-fire Debris Flow Triggering Concepts

The conceptual framework used in previous studies on post-fire debris flow triggering typically describes four principal categories: the terrain, fire alteration characteristics, the surface (or near surface), and rainfall [Negri, 2016; Staley et al., 2016, 2017]. Though these categories could be considered relatively broad, the combination of these factors statistically captures behavior of debris flow triggering across watersheds with heterogeneous scale and characteristics [Staley et al., 2017]. The goal of the input variables is to quantify the interaction between hydrologic and physical properties during debris flow initiation [Addison and Oommen, 2020; Rengers et al., 2020].

Utilized by the USGS/NOAA Debris flow hazard program, models from Staley et al., (2017) for the Southern California Region include; slopes $\geq 23^\circ$ with a high or moderate burn severity, burn severity measured as the proportion of the reflectance of certain wavelengths associated with vegetation before and after a fire per remote sensing [Addison and Oommen, 2020], and surface qualities measuring properties of the sediment such as clay content, grain size, and/or erodibility. Also included are rain measurements in amount and intensity of rainfall for differing storm intervals. Because the definition of a storm can change between studies, debris flow modeling has used rainfall intensity over time periods of 15, 30, and 60-minutes. Further studies suggest that short-intervals are best for the west coast due to the characteristic short and intense storms [Staley et al., 2017], and thus shorter periods will be focused upon in this study. While these measures are potentially non-exhaustive for any specific debris flow event, they allow quantifiable measures that are common across regions and models [Negri, 2016]. Other studies have also noted that factors such as aspect could also play a role in influencing debris flow initiation [Rengers et al., 2020].

3 Predicting Post-fire Debris Flow

Logistic regression is used in debris flow prediction to provide a statistical model that indicates if a watershed will trigger given; rainfall intensity, slope conditions, burned area, and sediment/surface properties. A logistic regression produces models that result in the probability between binary outcomes. The fundamental distinction between binary options will serve to

eliminate complexities introduced by partial or periodic failures that can be more commonly investigated for general landslide hazards.

The logit function is a curve defined by:

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

Where P is the probability ranging from 0 to 1 of the likelihood of occurrence and X is determined by the linear function:

$$X = \beta + C_1X_1 + C_2X_2 + C_3X_3 + \dots + C_nX_n \quad (2)$$

Where β and the C 's are empirically calculated coefficients that maximize the function with respect to the X 's, which are the input variables. For debris flows the linear function has been described specifically by the; terrain steepness (T), the wildfire severity (F), the properties of the sediment (S), and the rainfall intensity (R) that results in equation below;

$$X = f(T, F, S, R) \quad (3)$$

In this study we will define; T as proportion of upslope area burned at high or moderate severity with gradients $\geq 23^\circ$, F as differenced normalized burn ratio ($(dNBR)/1000$), S as either soil erodibility (KF) or shear wave velocity ($Vs30$), and R as 15-minute rainfall accumulation. We are introducing shear wave velocity for the first time in this study to evaluate how it will aid in the logistic predictive capacity. To ensure that the probability approaches zero when there is no rainfall, the rainfall intensity is used as a scalar for all the other independent variables including TR , FR , and SR . While there are studies that indicate no-rain failures can occur [Rengers et al., 2020], these events will not be examined as they are considered rare and possibly influenced by poorly defined factors. Rewriting equation 2 to reflect this specific application to post-fire debris flows:

$$X = \beta + C_1TR + C_2FR + C_3SR \quad (4)$$

The logistic regression considers parameters regardless of units of variables (X) that compose the coefficients (i.e., C). This allows the swapping of parameters for testing across varying units. The statistical spread of any one variable, however, can impact model stability.

4 Shear Stiffness in the Lab

To justify the use of shear wave velocity in a logistic model focused on debris flow initiation we examine the influence that shear resistance and thus shear wave velocity has on debris flow initiation in independent laboratory tests. Shear stiffness is the resistance of sediment to a shear

load. The shear stiffness is the initial shear modulus (G_o) which is the ratio of the shear stress (τ) to shear strain (γ) at the start of shear loading (Eq. 5). This is also equal to the product of the density of the sediment (ρ) and the squared shear wave velocity (V_s^2). The shear wave velocity is the speed at which a vertically propagating horizontal shear wave moves through a sample of sediment. Shear waves are body waves and the speed at which they travel through an arrangement of particles is a function of; particle density, bulk density, packing arrangement, number of particle contacts, ambient stress conditions, and other properties. Shear waves are not effectively transmitted through most fluids so the saturation level of the voids in the sediment sample has little impact on the shear wave velocity.

$$G_o = \frac{\tau}{\gamma} = \rho V_s^2 \quad (5)$$

Shear wave velocity has been used effectively as a predictor variable for seismic soil liquefaction because it captures the in-situ shear stiffness of the soil in question (e.g., [Kayen et al., 2013]) with shear loading being a primary mechanism for liquefaction failure. Low shear wave velocity values indicate that the soil is “loose” and has a high void ratio (e = volume voids / volume solids) with respect to critical void ratio (e_{crit}) and is therefore contractive and will likely generate positive excess pore pressures when sheared by incoming seismic waves. High shear wave velocity values indicate that the soil is “dense” and has a low void ratio with respect to the critical void ratio and is therefore dilative and will generate negative excess pore pressures when similarly loaded. The “loose” soil is more likely to fluidize when sufficiently loaded and results in the hazard of liquefaction, lateral spreading, flow failures, and other ground failures [Seed et al., 2003].

Seismic soil liquefaction triggering is a classification problem, akin to debris flow triggering, where there is a binary outcome. Post-event earthquake reconnaissance similarly strives to determine if YES the soil failed or NO it did not. These class models are then used to forecast the outcome for future events (e.g., [Moss et al., 2006; Yazdi and Moss, 2017], Smith and Moss, 2021). In concept the shear stiffness should then also provide some utility for debris flow triggering. The same particulate mechanics are in place, although the colluvial derived sediment of debris flows are typically more heterogeneous than river sands that comprise the bulk of the liquefaction hazards. The same physics-based concepts of soil state and pore pressure response control the behavior of debris flow sediments [Iverson et al., 1997; Moss, 2019]

Examining a controlled laboratory case, a case is further developed for the use of shear wave measurements to capture conditions for debris flow sediments. McKenna et al. (2012) performed a series of simple “dam break” experiments on sediment that they acquired from many slides and debris flows in the field. The experiment used a cylinder mold to contain a sediment sample, which was saturated from below, and then quickly released from its lateral confinement. This is identical to the analytical “dam break” problem as posed by [Hungr, 2008].

Plotting the shear rate collected in the beforementioned experiment versus the three modes of failure that the researchers used to bin the results, we see it provides a good predictor of response (Fig 1). A higher shear rate is indicative of a lower shear stiffness and hence a lower shear wave velocity.

Additional studies further suggest that a parameter such as the Shields stress, a function of material properties including shear stress and grain size that indicates the needed forces to initiate a flow, are essential for initiation in post-fire situations [Staley et al., 2017; Tillery and Rengers, 2019]. While this property is not as spatially available due to spatial heterogeneities of sediment properties, the inclusion of shear stress in its formulation further suggests a measure like shear wave velocity could be a reasonable predictor. Such findings taken in combination with the prior discussed shear rate studies suggest that a proxy for shear properties could allow better surface (S) property representation in models.

Figure 1. The effect of shear rate on samples of debris flow sediments of differing classifications of failure category derived from “dam break” experiments per McKenna et al. (2012).

5 Shear Stiffness in Hazard Assessment

Used commonly in seismic site response and liquefaction hazard assessment as mentioned above, the depth averaged V_{s30} represents the propagation and velocity of shear waves through the upper 30 m of surface material. It is measured in-situ through geophysical tests (Moss, 2008) for a given field site or through correlation with other spatially available information such as geology, topography, and steepness [Wald and Allen, 2007; Yong et al., 2016; Thompson, 2018].

Due to the seismic hazard application of $Vs30$ datasets, tectonically active areas like the western United States maintain high resolution spatial data mapping. Datasets are scaled at 1:24,000 and greater for global datasets [Thompson, 2018; Heath et al., 2020] . For regional assessment of debris flows, the 1:24,000 scale may offer reasonable resolution. This resolution also allows younger alluvial deposits to be classified by the surface slope, better capturing near surface characteristics in alluvial dominated watersheds compared to coarse datasets where steep terrain or rock outcrops could be filtered to reduce error [Yong, 2016; Yong et al, 2016].

6 Statistical Methodology

This study approaches debris flows as a binary class problem, between no flow and flow. As mentioned prior, partial slides are not considered to both simplify modeling approaches, and to also be more applicable towards immediate hazard assessment. Logistic regression has been used as a statistical method in the creation of prediction models for these binary flow outcome scenarios [Rupert et al., 2008; Negri, 2016; Staley et al., 2016, 2017; Rengers et al., 2020] and will be evaluated with metrics including; receiver operated characteristics (ROC) analysis, graphical output of multiple confusion matrices including Akaike information criterion (AIC) [Akaike, 1974], and Threat-score [Schaefer, 1990a]. These models were constructed with the KF for the surface property (S) and with the proposed $Vs30$ surface proxy for comparison across methods and datasets.

Data compiled by Staley et al. (2016) on over 900 occurrences of rainfall events resulting in both debris flow and no debris flow observations in the Southern California region was used for training and testing of the models. As noted earlier, we utilized the Proportion of upslope area burned at high or moderate severity with gradients $\geq 23^\circ$, differenced normalized burn ratio ((dNBR)/1000), Soil K-Factor (KF), and 15 min rainfall accumulation (R). This combination is identical to the “M1” model proposed by Staley et al. (2017) and utilized by the USGS for emergency assessment of debris flow risk. Incomplete entries within the dataset, such as those with no rainfall values, were omitted with marginal impact on model outputs resulting in 810 observations total in the dataset. This data exclusion was further justified by substituting missing values with near zero values ($\leq .001$) which offered marginal changes in model outcomes.

Figure 2. Map of general location of wildfires from 2003 to 2009 with debris flow observations in the Southern California Region used to create predictive models, compiled by Staley et al, 2017.

Vs30 information was collected across the same watersheds as those indicated in the Staley et al., (2016) dataset. Vs30 information is publicly available in spatial datasets hosted by the California Geologic Survey (CGS) and developed by the United States Geologic Survey [Thompson, 2018]. The GIS sampling of location specific Vs30 values for a given watershed coordinates from the dataset of debris flow characteristics was performed by referencing recorded coordinates of a watershed with the values indicated for the same location in the Vs30 dataset.

With Vs30 and KF information compiled across events, the central tendencies and dispersions were investigated for both model suitability and statistical utility. These measurements were separated by observed “Yes” and “No” responses for each variable with “Yes” being a debris flow had occurred and “No” being no debris flow had occurred. Visual inspection of the data trends also aided in determining if a transformation of the data should be implemented for a variable.

Logistic regression models were then created in a stepwise manner with increasing model inputs and complexity, starting with single variables. Next steps included scaling single variables by rainfall to produce the X terms in equation 4. Then these X terms were paired and assessed, followed by three term models to produce functions as shown in equations 6 and 7.

$$X = \beta + C_1(TR) + C_2(FR) + C_3(KF * R) \quad (6)$$

$$X = \beta + C_1(TR) + C_2(FR) + C_3(Vs30 * R) \quad (7)$$

Complementary to basic logistic regression methods are machine learning statistical approaches that can offer distinctions in new and existing model performance. This study also utilized a K-Fold cross-validation to confirm logistical regression results of KF and Vs30 models while also better accommodating the use of a single dataset for both “training” and “testing”. Metrics used between these models also evaluated ROC, AIC, and threat-score [Schaefer, 1990b].

7 Data Tendency and Spread

Box plots for the peak rainfall accumulation in millimeters within a 15-minute period (Acc15) is shown in Fig 3. The data collected for the slopes that did fail (YES) had both a higher central tendency and overall dispersion than the slopes that did not fail (NO). Similar box plots were observed for other loading variables such as, I15 (the peak rainfall intensity in mm/hr. that occurred within a 15 minute time period), PropHM23 (the proportion of the watershed area burned with slopes greater than 23 degrees), which can be found in Lyman (2021).

Figure 3. Boxplot distributions of 15-minute rainfall accumulation grouped by debris flow occurrence.

If, however, the surface resistance variable used in the “M1” model [Staley et al., 2017] we find that the predictive capacity is ambiguous. Fig. 4 shows the histogram and box plots for the variable KF (average erodibility index of the fine fragments of the sediment and other factors) as

defined by [Renard, 1997]. Here KF is plotted as natural log transform as the data is highly skewed, and this transform centralized the data which provides stability for statistical analysis [Sakia, 1992]. KF and $\ln(KF)$ provide limited predictive capacity for determining if a debris flow is likely or not. The central tendencies are nearly identical, and the dispersion are also relatively non informative as there is a non-unique value shared by both responses, making it ambiguous if a given variable will indicate a response in isolation.

Figure 4. Histograms and box-plots of KF and $\ln(KF)$ grouped by debris flow occurrence (Yes, No) in watersheds.

Looking at Vs30 both in a measured state and natural log transformed state displays similar issues of central tendency being relatively equivalent, with dispersion offering some reprieve as the interquartile range of “YES” is nearly twice as large as “NO” responses.

Figure 5. Histograms and box plots Vs30 and ln(Vs30) grouped by debris flow occurrence.

8 Logistic Regression Results

8.1 One Term Regression

Regression was first carried out using a single term to assess the relative contribution. The single term (Terrain, Surface, etc.) as a lone input with the rainfall (R) scalar, like pairings in eq 2, 5, and 6 are used during this processing. Scaling by R is necessary to avoid a prediction of debris flows when there is no rainfall loading. Each combination is referred in sequence of creation starting with X1, then X2, and onward. Lone X terms all show positive coefficients due to the scalar, with X1 showing the lowest AIC value of 684.6 indicating the most favorable model in the group, with X3 using KF showing the highest AIC of 721.5 indicating the inverse. The relatively poor evaluation of X3 with KF in this lone model starts to suggest that KF is a poor fit for the model and other variables could be used to improve overall model performance.

Table 1: Logistic regression coefficients, (standard error), and AIC values resulting from using X terms alone (eq.2) to predict debris flow response. The Surface variable is KF and constant is the β in equations 2, 5, and 6.

	Dependent variable:		
	Response		
	(1)	(2)	(3)
Terrain * Rain (X1)	0.784*** (0.066)		
Fire * Rain (X2)		1.300*** (0.120)	
Surface * Rain (X3)			1.951*** (0.178)
Constant	-2.983*** (0.195)	-2.725*** (0.176)	-3.040*** (0.203)
Observations	810	810	810
Akaike Inf. Crit.	684.620	705.235	721.513
Note:	* p<0.1; ** p<0.05; *** p<0.01		

The single term logistic regression results are plotted up as ROC curves and AUC values. All ROC curves placed well above the 1:1 (random) line and produced AUC values within 3% of each other (Figure 6). X1 as a function of T and R has the highest AUC at 83.1%. Focusing on just the X3 variable in Staley et al, 2017 ($X3 = KF \cdot R$) and as proposed here with Vs30 and natural logarithm transformed Vs30 (Figure 7) we see similar patterned ROC curve results. X3 as function of natural log transformed Vs30 producing the highest AUC value of 84.5% compared to the original KF AUC of 82.9%.

Figure 6. Receiver operating characteristic curve and AUC values resulting from using differing X terms (eq. 2) in a logistic regression to predict debris flow occurrence.

Figure 7. Receiver operating characteristic curve and AUC values resulting from using differing variations of the X3 term (the Surface Terms in eq. 2 and 4) in a logistic regression to predict debris flow occurrence.

8.2 Two Term Regressions

Seven pairs of X terms, including iterations with both KF and Vs30 with the X3 term, were then evaluated. AIC values ranged from nearly 648 for X1 and X2, to 668 for X1 and Vs30 X3. All combination's AIC results showed lower values than the best single term model with X1 producing an AIC of 685. While not directly comparable in isolation, the generally lower AIC values of two term models suggest improved performance from single terms models with considerably higher AIC values.

Table 2: Logistic regression coefficients, (standard error), and AIC values resulting from using X term pairs to predict debris flow response. Constant is equivalent to β in equations 2, 5, and 6.

	Dependent variable:						
	Response						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
X1	0.549*** (0.075)	0.566*** (0.078)		0.593*** (0.078)		0.527*** (0.079)	
X2	0.746*** (0.132)		0.910*** (0.131)		0.949*** (0.131)		0.800*** (0.142)
X3(KF)		0.982*** (0.212)	1.216*** (0.195)				
X3(Vs30)				0.0004*** (0.0001)	0.0005*** (0.0001)		
X3(ln[Vs30])						0.044*** (0.008)	0.050*** (0.008)
Constant	-3.347*** (0.216)	-3.429*** (0.229)	-3.449*** (0.227)	-3.393*** (0.227)	-3.447*** (0.227)	-3.616*** (0.240)	-3.515*** (0.235)
Observations	810	810	810	810	810	810	810
Akaike Inf. Crit.	648.163	664.631	662.941	667.903	665.265	653.839	664.530

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

ROC curves of these combinations produced overlapping curves with similar patterns at differing intervals of the plot. AUC values range between 84.5 % and 86.3%, with X1 and natural log transformed V30 with X3 having the largest value at 86.3%. The results of both AIC and AUC performing higher suggests a continuing improvement in model utility as complexity is added, along with increase model stability as seen by AUC values within a smaller range.

Figure 8. Receiver operating characteristic curve and AUC values resulting from using differing combinations of X terms in a logistic regression to predict if a debris flow occurrence.

8.3 Three Term Regressions

Now using three X terms to continue to add model complexity allowed the comparison of models with differing X3 terms. This provides a direct comparison to the model developed in Staley et al (2017). X1 and X2 terms were held constant across models. The 3 term regression results are shown in Table 3. Values derived from models including the new X3 term with Vs30 values are compared to the reference models containing X3 with Soil-KF regression (Staley et al., 2017). All models show similar values with natural log transformed Vs30 compared to the KF reference showing a slight difference. Vs30 AIC values did increase by 2.5 points for the natural log transformed Vs30 model, but such a difference may not be significant.

Table 3: Logistic regression coefficients, (standard error), and AIC values resulting from using three X terms per equations 5 and 6 to predict debris flow response. Constant is equivalent to β in equations 2, 5, and 6.

	<i>Dependent variable:</i>		
	Response		
	(1)	(2)	(3)
X1	0.421*** (0.085)	0.434*** (0.087)	0.437*** (0.084)
X2	0.652*** (0.136)	0.672*** (0.138)	0.564*** (0.146)
X3(KF)	0.693*** (0.216)		
X3(Vs30)		0.0003*** (0.0001)	
X3(ln[Vs30])			0.028*** (0.009)
Constant	-3.617*** (0.238)	-3.584*** (0.238)	-3.663*** (0.245)
Observations	810	810	810
Akaike Inf. Crit.	639.740	642.276	639.512
Note:	* p<0.1; ** p<0.05; *** p<0.01		

ROC characteristics, such as visible curve behavior, indicate marginal differences between the reference Soil-KF regression and both iterations of Vs30 values. AUC values also show similar near equivalency with only the AUC of non-log transformed Vs30 showing a 0.1 % decrease compared to other regressions. Despite the exceedingly similar results, logarithm transformed Vs30 does indicate the best performance of the three models tested by 0.2 points in AIC, and 0.2% in AUC difference from the next best model.

Figure 9. Receiver operating characteristic curve and AUC values resulting from using three X terms (eq. 5 and 6) in a logistic regression to predict debris flow occurrence.

9 K-fold Cross Validation Results

Noting the similarity in results from three X term models in a typical logistic regression, additional focus on how training and testing datasets are divided given a similar dataset is warranted. Without this consideration, differing results may be reached within the same dataset. K-fold cross validation with 10 folds allows such splitting in a more robust and objective manner. Coefficient and intercept values from basic logistic regression result in identical data trends from normal regression, with the natural log transformed dataset offering the lowest AIC value. The near equivalency to models produced without this cross validation indicate that for this dataset the models developed through logistical regression are sufficient.

Table 4: Logistic regression with 10-fold cross validation coefficients, (standard error), and AIC values resulting from using three X terms to predict debris flow response. Constant is equivalent to β in equations 2, 5, and 6.

	Dependent variable:		
	Response		
	(1)	(2)	(3)
X1	0.421*** (0.085)	0.434*** (0.087)	0.437*** (0.084)
X2	0.652*** (0.136)	0.672*** (0.138)	0.564*** (0.146)
X3(KF)	0.693*** (0.216)		
X3(Vs30)		0.0003*** (0.0001)	
X3(ln[Vs30])			0.028*** (0.009)
Constant	-3.617*** (0.238)	-3.584*** (0.238)	-3.663*** (0.245)
Observations	810	810	810
Akaike Inf. Crit.	639.740	642.276	639.512
Note:	* p<0.1; ** p<0.05; *** p<0.01		

Figure 10. Receiver operating characteristic curve and AUC values resulting from using three X terms (eq. 5 and 6) in a 10-fold cross validated logistic regression to predict if a debris flow occurrence.

Additionally, similar to more basic logistical regression, the ROC patterns are largely similar between differing variables. AUC is again 0.2% higher when utilizing the natural log transformed Vs30 data. These values are equivalent to those produced without cross-validation testing, implying models created should represent the best performance model given the inputs and the dataset.

Table 6: Threat scores resulting from logistic regression models with 10-fold, k-fold cross validation calculated per equation 7 that compares true-positive results to false-positives, and false-negatives. A score of 1 represents a perfect model.

	KF	Vs30	Ln(Vs30)
Threat Score	0.41	0.41	0.39

The nature of K-fold further allows for a threat score to be calculated for a single dataset, as the same dataset is both “training” and “testing” data for the resultant model. Threat score (Table 1) is another measure of performance where each incorrect prediction (i.e. false-positive or false-negative) reduces the score, and a perfect model has a value of 1. Threat scores resulted in KF and Vs30 producing the same score of 0.41, while natural logarithm transformed Vs30 scored a 0.39. For this metric KF and Vs30 are near equivalents, with both offering higher predictive performance compared to natural log transformed Vs30.

10 Discussion

The detailed statistical interrogation of the terms and input variables for post-fire debris flow triggering for this Southern California dataset, and the premise that a variable better representing the properties of the soil in shear may better aid in forecasting debris flows, show improved results over previously used models. However, the results beyond statistical measures in the context of the variables being fielded and what new variables represent could be just as significant in consideration for reasons including; future model application/development and hazards such as potential spurious correlations.

Such context includes noting that Soil KF represents a relative measure of the long-term susceptibility of soil to erode in sheet and rill erosion by taking into account soil texture, organic matter, hydraulic conductivity, and other soil measures [Renard, 1997]. KF could then be useful in large watershed and agricultural scales as it was originally designed for, yet this may not be the case at debris-flow watershed scales both in this study and in other post-fire watersheds. In the dataset used in this study, the KF was extracted from a national 1:250,000 scale database hosted by the United States Department of Agriculture [Schwarz and Alexander, 1995]. Such a scale is similar in scale to a USGS map for the entirety of the United States of America, potentially resulting in large spatial errors for the average 1 km² contributing areas found in this study. Such a scale could offer the benefit of data availability across vast distances of the United States, including state or larger analyses. There is also the possible issue with the use of this variable with the quick and distinct changes in soil, vegetation, and hydrologic properties after a fire as noted earlier. There could be the possibility of higher resolution data in certain localities but for general application this study does not by in large consider them. These challenges mixed with the statistical concerns in distribution and modeling leave room for other measures.

Shear wave velocity as a measure of the shear stiffness of soil initially appears as a good measure of the likelihood for triggering as it has been applied in related fields and topics. While local field measurements of debris flow sources were not used, the depth-average 30 m shear wave velocity (V_{s30}) has been mapped from both in situ measurements and topographic correlations, and is used as a proxy for seismic site conditions. Data in this study is hosted by the USGS and offered the implementation of slope and topography proxy V_{s30} at a 1:24,000 scale for the state of California [Thompson, 2018]. This resolution could also adversely impact small watersheds, and locations with highly variable/complex surface characteristics between each other. The resolution seen in California may also not extend to other locations, but databases do exist that offer global coverage like those by [Heath et al., 2020] that offer improved resolution compared to KF. Variability at depth is also a concern, as the upper 30 m of material likely blurs lithologies at depth, as it can reasonably be expected that near surface characteristics differ at greater depths. Further considering that shear wave velocity is a more fundamental measurement compared to KF, it could stand to be a more applicable variable for debris-flow hazard assessment through potential spatial transferability, and ease of use being a fundamental, or non-derived, property of soil. Some issues with V_{s30} could also be alleviated through the collection of field measurements through geophysical testing in a sufficient number that could lend to the creation of a more suitable property to implement in hazard models. There is also the possibility that correlation of field collected velocities could be used in the creation of a transfer function to a more spatially available dataset, such as V_{s30} , that could permit the use of a V_{s30} model in differing regions.

Looking at the differences between KF and V_{s30} as part of the tested predictive models, the results are largely similar. With near equivalent statistical performance for some uses, the ability to use a model in practice may then offer another view of evaluating predictions. As such, a variable that is easier to manage or quicker to obtain could be more effective in practice. Similarly, a simpler or more streamlined model may permit speed of application in emergency situations where resources might be limited. A case where this may be applicable is looking at the similarity in results from two terms models, of which X_1 and either X_3 or $X_3(\text{LN}[V_{s30}])$ has the performance values similar to the best three term models. The use of only two X terms could reduce the number of input variables and increase data collection speed, and thus improve implementation. Such speed could offer emergency modeling for recently burned areas a distinct advantage with potentially little to no loss in model performance depending on acceptable risk acceptance. Such removal of the F factor in the two term model mentioned prior measured by the normalized burn ratio may be moot when considering its measurement is used in identifying burned watersheds prior to a debris flow hazard assessment [Key and Benson, 2005; Addison and Oommen, 2020] suggesting context should be carefully evaluated in such cases.

The method in splitting this dataset could also bring with it important considerations when discerning if V_{s30} is a better variable compared to KF. Per the methodology established earlier, all models used the entirety of the dataset to create a model using logistic regression. While using all entries in a dataset can help to generalize and thus make it more applicable to varying situations, it can also produce overly optimistic models. This implies that models that did not use cross-validation should be validated against additional data to assure performance. K-fold cross validation relieves some of these concerns as it validates by splitting the existing dataset multiple

times [Cawley and Talbot, 2010]. Additional datasets can reduce issues that may arise in splitting datasets, but cross-validation offers an avenue for when this is not possible.

Statistical models also have a plethora of metrics to assess performance to observed outcomes and between models directly. While the metrics used in this study are by no means exhaustive, they are commonly accepted in statistical modeling and in previous work in debris flow [Rupert et al., 2008; Negri, 2016; Staley et al., 2016, 2017; Addison and Oommen, 2020]. Works by [Rengers et al., 2020] suggest additional metrics could be of use in certain areas but produce values with comparable utility as the ones provided in this study. This further implies context is important, or at least part of a toolset, to assess the physical implications of statistical model inputs.

Considering the above context, Vs30 as an input parameter offers an improvement in most measures of modeling outcomes in the analyzed dataset compared to KF, both through objective and subjective measures. The fundamental nature of shear stiffness suggests that capturing the underlying physical properties of the soil is a productive avenue for further investigation. Such work can implement further statical analysis, lab and field-based experiments, and/or integration into physical aspects including volume or runout prediction. Similarly, additional data and associated resolution could then offer affirmation in substituting KF with Vs30, or another path to the underlying mechanics of debris flow initiation to be captured.

11 Conclusions

This study looks at different input variables for forecasting post-fire debris flow triggering. A proxy of the shear stiffness of soil (Vs30) is compared to the current property of soil erodibility (KF). Logistic regression models of debris flow probability using Vs30 rather than KF show slight improvement in predicting outcomes for Southern-California datasets. There are also slight improvements shown using the machine learning statistical approach of K-Fold cross validation. It was found that two term models produced similar performance to three term models which could offer a reasonable substitution in cases where data procurement and processing time is important, such as in emergency assessment. Further improvement could be found with additional or refined surface variables. One way of achieving this may be through the creation of a transfer function of field measures of shear wave velocity correlated with more spatially available datasets such as Vs30.

Acknowledgements

Integral to the success of this work was the generous support of the Warren J. Baker Endowment in Project-Based Learning and Robert D. Koob Endowment for Student Success. Additional resources were offered by the California Polytechnic State University – San Luis Obispo Graduate Assistantship Fund. We also want to thank research at the USGS, NOAA, and the California Geologic Survey for their data collection and there works of debris flows. Finally, we would like to thank Dennis Staley, Jason Kean, and the many other collaborators for their work and models that this work builds upon.

526

527 **References**

- 528 Addison, P., and T. Oommen (2020), Post-fire debris flow modeling analyses: case study of the
529 post-Thomas Fire event in California, *Nat. Hazards*, 100(1), 329–343, doi:10.1007/s11069-
530 019-03814-x.
- 531 Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*,
532 19(6), 716–723.
- 533 Cannon, S. H., and J. DeGraff (2009), The increasing wildfire and post-fire debris-flow threat in
534 western USA, and implications for consequences of climate change, in *Landslides—disaster
535 risk reduction*, pp. 177–190, Springer.
- 536 Cawley, G. C., and N. L. C. Talbot (2010), On over-fitting in model selection and subsequent
537 selection bias in performance evaluation, *J. Mach. Learn. Res.*, 11, 2079–2107.
- 538 Dowling, C. A., and P. M. Santi (2014), Debris flows and their toll on human life: a global analysis
539 of debris-flow fatalities from 1950 to 2011, *Nat. hazards*, 71(1), 203–227.
- 540 De Graff, J. V (2018), A rationale for effective post-fire debris flow mitigation within forested
541 terrain, *Geoenvironmental Disasters*, 5(1), 1–9.
- 542 Gartner, J. E., S. H. Cannon, and P. M. Santi (2014), Empirical models for predicting volumes of
543 sediment deposited by debris flows and sediment-laden floods in the transverse ranges of
544 southern California, *Eng. Geol.*, 176, 45–56, doi:10.1016/j.enggeo.2014.04.008.
- 545 Heath, D. C., D. J. Wald, C. B. Worden, E. M. Thompson, and G. M. Smoczyk (2020), A global
546 hybrid VS 30 map with a topographic slope–based default and regional map insets, *Earthq.
547 Spectra*, 8755293020911137.
- 548 Hungr, O. (2008), Simplified models of spreading flow of dry granular material, *Can. Geotech. J.*,
549 45(8), 1156–1168.
- 550 Iverson, R. M., M. E. Reid, and R. G. LaHusen (1997), Debris-flow mobilization from landslides,
551 *Annu. Rev. Earth Planet. Sci.*, 25, 85–138.
- 552 Kayen, R., R. E. S. Moss, E. M. Thompson, R. B. Seed, K. O. Cetin, A. Der Kiureghian, Y. Tanaka,
553 and K. Tokimatsu (2013), Shear-wave velocity–based probabilistic and deterministic
554 assessment of seismic soil liquefaction potential, *J. Geotech. Geoenvironmental Eng.*,
555 139(3), 407–419.
- 556 Kean, J. W., D. M. Staley, J. T. Lancaster, F. K. Rengers, B. J. Swanson, J. A. Coe, J. L. Hernandez,
557 A. J. Sigman, K. E. Allstadt, and D. N. Lindsay (2019), Inundation, flow dynamics, and
558 damage in the 9 January 2018 Montecito debris-flow event, California, USA: Opportunities
559 and challenges for post-wildfire risk assessment, *Geosphere*, 15(4), 1140–1163,
560 doi:10.1130/ges02048.1.
- 561 Key, C., and N. Benson (2005), Landscape Assessment (LA) Sampling and Analysis Methods, p.
562 LA 1-51.
- 563 Kotok, E. I. (1935), Kotok and Kraebel on Flood and Erosion Problems, *Trans. Am. Soc. Civ. Eng.*,
564 1935, 1350–1355.
- 565 Lyman (2020). [Incorporating Shear Resistance into Debris Flow Triggering Model Statistics](https://digitalcommons.calpoly.edu/theses/2254/).
566 Thesis in partial fulfilment of MS degree. California Polytechnic State University, San Luis
567 Obispo. <https://digitalcommons.calpoly.edu/theses/2254/>

- McGuire, L. A., F. K. Rengers, J. W. Kean, D. M. Staley, and B. B. Mirus (2018), Incorporating spatially heterogeneous infiltration capacity into hydrologic models with applications for simulating post-wildfire debris flow initiation, *Hydrol. Process.*, 32(9), 1173–1187.
- McKenna, J. P., Santi, P. M., Amblard, X., & Negri, J. (2012). Effects of soil-engineering properties on the failure mode of shallow landslides. *Landslides*, 9(2), 215–228.
- Moss, R. E. S. (2008). Quantifying measurement uncertainty of thirty-meter shear-wave velocity. *Bulletin of the Seismological Society of America*, 98(3), 1399–1411.
- Moss, R. E., R. B. Seed, R. E. Kayen, J. P. Stewart, A. Der Kiureghian, and K. O. Cetin (2006), CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential, *J. Geotech. Geoenvironmental Eng.*, 132(8), 1032–1051.
- Moss, R. E. S. (2019), Commonalities between debris flows and flow failures, in *Proceedings of the Seventh International Conference on Debris-flow Hazards Mitigation, Golden, Colorado, USA, June 10-13, 2019*, vol. 1, Colorado School of Mines. Arthur Lakes Library.
- Negri, J. A. (2016), Evaluation and Validation of Multiple Predictive Models Applied To Post-Wildfire Debris-Flow Hazards, 2016 - *Mines Theses Diss.*, 1, doi:10.1017/CBO9781107415324.004.
- Renard, K. G. (1997), *Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE)*, United States Government Printing.
- Rengers, F. K., L. A. McGuire, N. S. Oakley, J. W. Kean, D. M. Staley, and H. Tang (2020), Landslides after wildfire: initiation, magnitude, and mobility, *Landslides*, (December 2019), doi:10.1007/s10346-020-01506-3.
- Rowe, P. B., C. M. Countryman, and H. C. Storey (1949), Probable peak discharges and erosion rates from southern California watersheds as influenced by fire, *US Dep. Agric. For. Serv. Calif. For. Range Exp. Station*. 107 p.
- Rupert, M. G., S. H. Cannon, J. E. Gartner, J. A. Michael, and D. R. Helsel (2008), *Using Logistic Regression to Predict the Probability of Debris Flows in Areas Burned by Wildfires, Southern California, 2003-2006*.
- Sakia, R. M. (1992), The Box-Cox transformation technique: a review, *J. R. Stat. Soc. Ser. D (The Stat.)*, 41(2), 169–178.
- Schaefer, J. T. (1990a), The critical success index as an indicator of warning skill, *Weather Forecast.*, 5(4), 570–575.
- Schaefer, J. T. (1990b), The critical success index as an indicator of warning skill, *Weather Forecast.*, 5(4), 570–575.
- Schwarz, G. E., and R. B. Alexander (1995), *State soil geographic (STATSGO) data base for the conterminous United States*.
- Seed, R. B., K. O. Cetin, R. E. S. Moss, A. M. Kammerer, J. Wu, J. M. Pestana, M. F. Riemer, R. B. Sancio, J. D. Bray, and R. E. Kayen (2003), Recent advances in soil liquefaction engineering: a unified and consistent framework, in *Proceedings of the 26th Annual ASCE Los Angeles Geotechnical Spring Seminar: Long Beach, CA*.
- Schmidt, J., & Moss, R. (2021). Bayesian hierarchical and measurement uncertainty model building for liquefaction triggering assessment. *Computers and Geotechnics*, 132, 103963.
- Staley, D. M., J. A. Negri, J. W. Kean, J. L. Laber, A. C. Tillery, and A. M. Youberg (2016), Updated logistic regression equations for the calculation of post-fire debris-flow likelihood in the western United States, , (2016–1106), doi:10.3133/ofr20161106.

- Staley, D. M., J. A. Negri, J. W. Kean, J. L. Laber, A. C. Tillery, and A. M. Youberg (2017), Prediction of spatially explicit rainfall intensity–duration thresholds for post-fire debris-flow generation in the western United States, *Geomorphology*, 278, 149–162, doi:10.1016/j.geomorph.2016.10.019.
- Thompson, E. M. M. (2018), An Updated Vs30 Map for California with Geologic and Topographic Constraints: U.S. Geological Survey data release, *US Geol. Surv. Data Release*, doi, 10, F7JQ108S, doi:/10.5066/F7JQ108S.
- Tillery, A. C., and F. K. Rengers (2019), Controls on debris-flow initiation on burned and unburned hillslopes during an exceptional rainstorm in southern New Mexico, USA, *Earth Surf. Process. Landforms*, n/a(n/a), doi:10.1002/esp.4761.
- USGS (2005), NOAA-USGS Debris-Flow Warning System — Final Report, *US Geol. Surv. Circ.*, 44.
- Wald, D. J., & Allen, T. I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, 97(5), 1379-1395.
- Westerling, A. L., H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam (2006), Warming and earlier spring increase western US forest wildfire activity, *Science* (80-.), 313(5789), 940–943.
- Yazdi, J. S., and R. E. S. Moss (2017), Nonparametric liquefaction triggering and postliquefaction deformations, *J. Geotech. Geoenvironmental Eng.*, 143(3), 4016105.
- Yong, A. (2016), Comparison of measured and proxy-based VS 30 values in California, *Earthq. Spectra*, 32(1), 171–192.
- Yong, A., E. M. Thompson, D. J. Wald, K. L. Knudsen, J. K. Odum, W. J. Stephenson, and S. Haefner (2016), *Compilation of V S30 data for the United States*, US Geological Survey.

635   **Statements and Declarations**

636   Funding

637   Financial support came from the Warren J. Baker Endowment in Project-Based Learning and
638   Robert D. Koob Endowment for Student Success at Cal Poly. Additional resources were offered
639   by the California Polytechnic State University – San Luis Obispo Graduate Assistantship Fund.

640   Competing Interests

641   The authors have no relevant financial or non-financial interests to disclose.
642

643   Authors Contributions

644
645   Both authors contributed to the completion of this research. Robb Moss conceived of the idea
646   and developed the proof of concept, including initial statistical analysis and modeling. Noah
647   Lyman performed data collection and thorough statistical analysis to fill out the study.
648

649   Portions of this manuscript were taken from the MS Thesis of Noah Lyman, which was edited
650   and commented on by Robb Moss. The manuscript as a whole was drafted by Robb Moss with
651   edits, comments, and contributions by Noah Lyman.
652

653   Both authors approve of the final manuscript.
654