Runout model evaluation based on back-calculation of building damage

Katherine R. Barnhart[1*] and *Jason W.* Kean[1]

[1] U.S. Geological Survey, Box 25046 DFC MS 966; Denver, CO 80225-0046, USA

Abstract. We evaluated the ability of three debris-flow runout models (RAMMS, FLO2D and D-Claw) to predict the number of damaged buildings in simulations of the 9 January 2019 Montecito, California, debris-flow event. Observations of building damage after the event were combined with OpenStreetMap building footprints to construct a database of all potentially impacted buildings. At the estimated event volume, all models overpredict the number of damaged buildings by a factor of 1.5–3.

1 Introduction

Debris flows pose a hazard to buildings located within the extent of inundated area [1]. The force exerted by a debris flow on a structure can result in damage ranging from slight (e.g., failure of non-load bearing components) to complete destruction (e.g., substantial structural damage, removed from foundation). A reliable method for forecasting building damage in an anticipated debris-flow runout zone would be useful for decision making activities such as evacuation planning.

A fragility function provides the link between hazard intensity and the corresponding likelihood of damage. Hazard intensity refers to a physical characteristic of the hazard that can reliably predict the likelihood of damage, such as debris-flow depth [2], the ratio of debris-flow depth to building height [3], and the product of debris-flow depth and velocity squared [1].

Presuming a reliable method to forecast hazard intensity, a fragility function may be used to provide a vulnerability assessment of one or more buildings. Accordingly, forensic evaluation of past events in which a group of buildings experienced damage provides an opportunity to evaluate the coupling of fragility functions to models used to forecast hazard intensity.

We evaluated the ability of three runout models to predict building damage by coupling the output of each runout model to a previously defined fragility function. We were interested in the sensitivity of predicted building damage to runout model choice, event volume, and flow mobility. Consequently, we used runout model simulations initialized with a range of event volumes and mobility values. Because our results document the relative role of runout model, event volume, and flow mobility on predicting building damage, they inform which areas of study may be most fruitful for reducing uncertainty in forecasts of building damage.

Our study was done in the context of the 9 January 2018 Montecito, California, debris-flow event (hereafter "Montecito event") [4–6]. This event occurred after intense rain (5-minute intensity of 157 mm/hr) fell on the recently burned Santa Ynez Mountains. The event mobilized sediment from hillslopes and channels [7,8] into a boulder-laden slurry that ran out onto a ~4 km-wide coalesced alluvial fan located between the Santa Ynez Mountains and the Pacific Ocean. The debris-flow runout inundated a combined area of 2.6 km² and resulted in 23 fatalities, at least 167 injuries, and many damaged homes [4–6].

2 Methods and Data

Fragility functions were fit by Kean et al. [4] for wood-framed buildings using observed debris-flow depth, inferred momentum flux, and observed damage. For simplicity, and because it is most directly related to the observational data, here we only evaluated back-calculated building damage based on maximum simulated debris-flow depth. For each of the simulations presented in Barnhart et al. [9], we extracted the maximum simulated debris-flow depth at every considered building and classified the building into a predicted damage state. Use of maximum simulated debris-flow depth as our measure of hazard intensity assumes that instantaneous depth is a reliable measure of the depth effective for damaging buildings. We evaluated model validity using the frequency bias, a standard metric in binary classification that is calculated as the ratio of the number of buildings with predicted damage and the number of buildings with observed damage. The frequency bias was chosen as an evaluation metric because it evaluates whether the correct number of damaged buildings was forecast.

Our analysis required a dataset of buildings, established fragility functions, and simulation results. Additionally, we interpreted the results in the context of estimated event size.

---

* Corresponding author: krbarnhart@usgs.gov
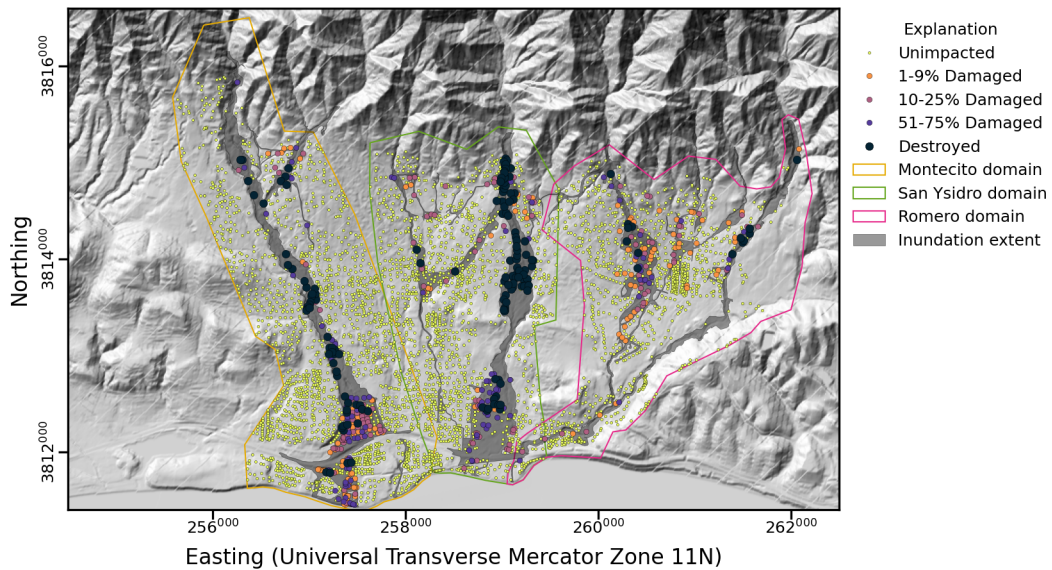
Fig. 1. Map of building damage incurred during the 9 January 2018, Montecito debris-flow event. All buildings from the California Department of Forestry and Fire Protection (CAL FIRE) database and OpenStreetMap located within the simulation domains are depicted as dots with color indicating damage state.

2.1 Building dataset

After the event, building inspectors produced a database of damaged homes that was published by the U.S. Geological Survey [10]. For this event, the California Department of Forestry and Fire Protection (CAL FIRE) building inspectors classified impacted buildings into four ordinal damage state categories: 1%–9% damaged, 10%–25% damaged, 51%–75% damaged, and destroyed. We supplemented this database with the location of all buildings in the considered simulation domains from OpenStreetMap (OSM, https://www.openstreetmap.org/, database accessed November 12, 2021) (Fig. 1). We de-duplicated the building dataset by removing OSM-sourced buildings with building footprints that overlapped with a building in the CAL FIRE dataset. The OSM-sourced buildings were categorized as unimpacted, yielding a total of five building damage categories. The final building dataset contained 4002 unimpacted buildings, 127 buildings with 1%–9% damage, 126 buildings with 10%–25% damage, 114 with buildings 51%–75% damage, and 162 destroyed buildings.

2.2 Fragility functions

Kean et al. [4] estimated the flow depth (h, meters) at each damaged building within the CAL FIRE building damage dataset and fit a set of fragility functions for wood-framed buildings. These fragility functions have the form:

$$P^i(h) = \Phi\left[\frac{1}{\beta_i}\ln\left(\frac{h}{\overline{h_i}}\right)\right] \qquad (1)$$

Where P^i is the probability of reaching or exceeding damage state i, Φ is the standard normal cumulative distribution function, β_i is a parameter indicating the uncertainty in the fragility function and was fit based on observation data, and $\overline{h_i}$ is the median observed debris-flow depth in damage state i. Kean et al. [4] report fragility functions for both h and momentum flux.
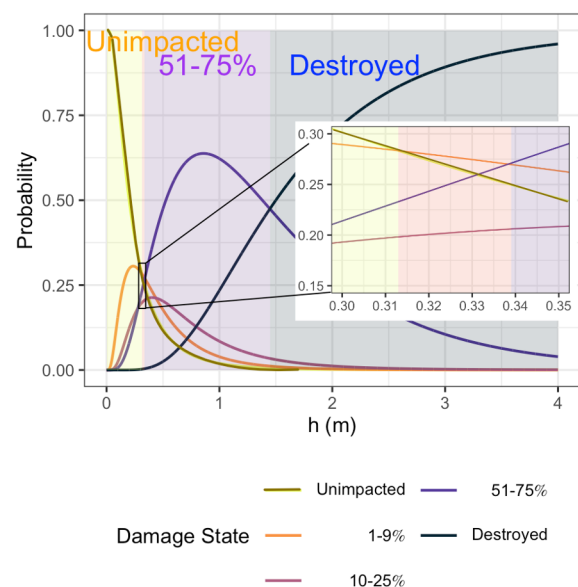

Fig. 2. Probability of damage state as a function of inundation depth, h. The 1–9% damage category is most probable for depths between 0.31 and 0.34 m (inset).

We classify the value of simulated maximum debris-flow depth at each building into predicted damage states by choosing the most probable damage state for a given debris-flow depth (Fig. 2). The values of $\overline{h_i}$ and β_i fit from the Montecito event result in classification into only four categories: unimpacted, 1–9% damaged, 51–75% damaged, and destroyed. This occurs because there is no value of h for which the probability of 10–25% damage is the largest.

2.3 Simulated flow depth

We used simulation results from a prior study [9] that evaluated three different runout models (RAMMS [11], FLO2D [12], and D-Claw [13,14]) across three different domains of the Montecito event runout paths
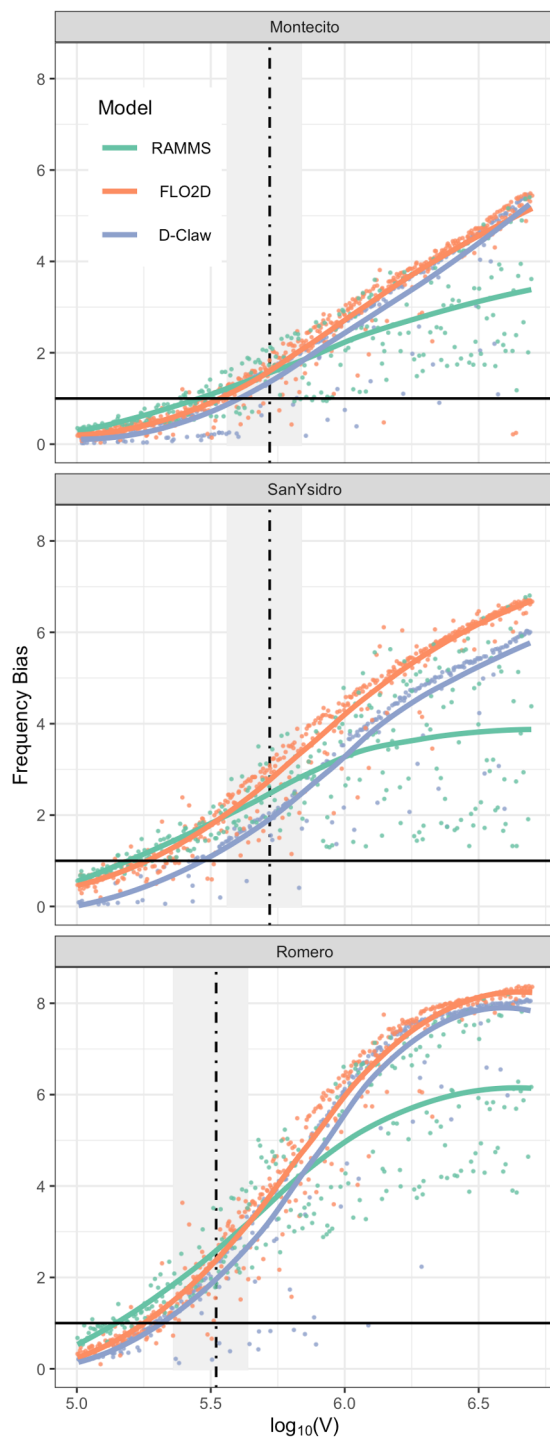
Fig. 3. Frequency bias as a function of the $\log_{10}$ of volume, V. Rows indicate different model domains, and color indicates model used. The vertical dashed line indicates the estimate of observed volume for that domain [9], and the horizontal solid line indicates a value of frequency bias equal to 1, or perfect prediction. Gray vertical rectangles depict an arbitrary 70%–130% range. Each simulation is depicted as a dot and the solid line reflects the conditional mean for each model.

(Montecito, San Ysidro, and Romero domains depicted in Fig. 1). In the prior study, the authors used a Latin hypercube sampling study to simulate debris-flow runout at Montecito under a range of debris-flow volumes, V (m³), and material properties. Each model uses a different set of governing equations and, thus, a different set of inputs that describe the mobility of

debris-flow material. For a given model and domain, the number of simulations was determined as 100× the number of model free parameters, N_p, ($N_p = 3, 5,$ and 4 for RAMMS, FLO2D, and D-Claw, respectively, as described by Barnhart et al. [9]). For each simulation, the maximum debris-flow depth was recorded at each grid cell (5-m cell sides).

2.4 Observed event size and uncertainty

Prior work estimated the total amount of sediment deposited in the event [4], eroded from the hillslopes [7], and eroded from the channels [8]. Barnhart et al. [9] combined the sediment volumes upstream from the three domains with an estimate of water volume based on rainfall-runoff analysis to produce an estimate of total event volumes for each domain: 531,000 m³ ($\log_{10}(V) = 5.73$) for Montecito, 522,000 m³ ($\log_{10}(V) = 5.73$) for San Ysidro, and 332,000 m³ ($\log_{10}(V)=5.52$) for Romero. Barnhart et al. [9] considered an arbitrary factor of two uncertainty estimate for the event volume (50%–200%). Because more recent work estimating the erosion from hillslopes and channels matches well with the estimates of deposit volume, here we consider a smaller, though still arbitrary, uncertainty range of 70%–130% on event volume (depicted in Fig. 3).

3 Results

Our analysis results in a relation between frequency bias and simulated volume for each model and domain (Fig. 3). The frequency bias is the ratio of the number of predicted and observed damaged buildings. Thus, the frequency bias reflects how well simulations are able to forecast the number of damaged buildings. It is equal to 1.0 when the observed and predicted number of damaged buildings is the same; a value less than 1.0 occurs when the predicted number of damaged buildings is smaller than observed; and a value greater than 1.0 occurs when the predicted number of buildings is greater than observed.

For ease of interpreting the relative influence of runout model and event volume on performance, we calculated the conditional mean of the frequency bias as a function of volume—that is, for each site and model, we calculated the running mean as a function of the flow volume using a LOESS fit. Frequency bias increases with increasing volume for all models and all domains, reflecting a higher number of damaged buildings for larger event volumes. The relation between frequency bias and volume varies depending on model and domain. Models are in closer agreement, except at the highest volumes, in the Montecito domain. Models have the least agreement for the San Ysidro domain. Scatter is considerable in the frequency bias predicted at most volumes, and this scatter reflects simulations with similar volumes but different flow mobility input parameters. Scatter is generally higher for larger volumes.

At the event volume estimated for each domain (vertical dashed line in Fig. 3), the local mean of the frequency bias is larger than 1.0 for all models and all

Table 1. Conditional mean of frequency bias for each model and domain at the observed event volume.

	Montecito	San Ysidro	Romero
RAMMS	1.5	2.4	2.6
FLO2D	1.6	2.7	2.4
D-Claw	1.4	1.9	2.0

domains (Table 1). For each domain, D-Claw has the lowest frequency bias at most volumes. Above ~10^6 m³, RAMMS has the lowest frequency bias. For the Montecito domain, all models produce a frequency bias of 1.0 when the simulated volume is on the lower end of the event size uncertainty range. The simulated volume needed to produce a frequency bias of 1.0 ranges from 30%–76% of the estimated volume (Table 2).

4 Discussion and conclusions

At the estimated event volume, all models overpredict the number of damaged buildings by 1.5–3. This result indicates that even if the volume of the event had been known in advance, use of a runout model and debris-flow depth-based fragility functions would have overestimated the total number of damaged buildings.

The fragility functions indicate a 0.32-m maximum flow depth for the transition between an undamaged and damaged wood frame buildings (Fig. 2). For events impacting building types with different strength properties than wood-framed construction, the relevant flow depth threshold may be different.

Coupling the runout models with the fragility functions and evaluating model performance based on the number of damaged buildings is essentially a test of whether the runout model can correctly predict the aerial extent impacted by depths thicker this threshold. However, many runout models struggle to reliably simulate the extent impacted by flows in this depth range—likely because of the absence of a mechanism to create levees. A standard practice to address the overprediction of debris-flow extent is to use a threshold ranging between 0.1 and 0.5 m to extract debris-flow extent from simulated maximum flow depth [9,15,16]. The similar magnitude of the fragility function threshold (0.3 m) and the threshold commonly used to interpret debris-flow extent from simulated maximum flow depth indicates that improving representation of flow edges in runout models may have the largest effect on improving building damage forecasts.

Pre-event estimates of post-fire debris-flow volume are themselves uncertain [17]. The best-studied relation between rainfall intensity and mobilized debris-flow sediment comes from southern California and caries an order of magnitude prediction uncertainty range. Interpretation of pre-event vulnerability assessment for building damage consequently must reflect the uncertainty in forecast rainfall, debris-flow runout conditional on rainfall, and building damage conditional on runout.

In the context of post-fire debris flows, where evacuation fatigue is a concern (residents in areas susceptible to post-fire debris flows may have only

Table 2. Proportion of observed volume required to produce a frequency bias of 1.0.

	Montecito	San Ysidro	Romero
RAMMS	0.54	0.30	0.43
FLO2D	0.55	0.35	0.55
D-Claw	0.76	0.57	0.64

recently returned to their homes after evacuation during a fire), downsides to overprediction of building damage may be substantial. Discussion and communication with the local emergency management community may elucidate whether a 1.5–3 estimate is usable or if more accurate forecasts of building damage are needed.

References

1. M. Jakob, D. Stein, and M. Ulmi, Nat. Hazards **60**, 241 (2012)
2. S. Fuchs, K. Heiss, and J. Hübl, Nat. Hazards Earth Syst. Sci. **7**, 495 (2007)
3. R. Totschnig, W. Sedlacek, and S. Fuchs, Nat. Hazards **58**, 681 (2011)
4. J. W. Kean, D. M. Staley, J. T. Lancaster, F. K. Rengers, et al., Geosphere **15**, 1140 (2019)
5. J. T. Lancaster, B. J. Swanson, S. G. Lukashov, et al., Environ. Eng. Geosci. **27**, 3 (2021)
6. N. S. Oakley, F. Cannon, R. Munroe, J. T. Lancaster, et al., Nat. Hazards Earth Syst. Sci. **18**, 3037 (2018)
7. P. Alessio, T. Dunne, and K. Morell, J. Geophys. Res. Earth Surf. **126**, e2021JF006108 (2021)
8. K. D. Morell, P. Alessio, T. Dunne, and E. Keller, Geophys. Res. Lett. **48**, e2021GL095549 (2021)
9. K. R. Barnhart, R. P. Jones, D. L. George, B. W. McArdell, et al., J. Geophys. Res. Earth Surf. **126**, e2021JF006245 (2021)
10. J. W. Kean, D. M. Staley, J. T. Lancaster, F. K. Rengers, et al., US Geol. Surv. Data Release (2019) http://doi.org/10.5066/P9JQJU0E
11. M. Christen, J. Kowalski, and P. Bartelt, Cold Reg. Sci. Technol. **63**, 1 (2010)
12. J. S. O'Brien, P. Y. Julien, and W. T. Fullerton, J. Hydraul. Eng. **119**, 244 (1993)
13. D. L. George and R. M. Iverson, Proc. R. Soc. Math. Phys. Eng. Sci. **470**, 20130820 (2014)
14. R. M. Iverson and D. L. George, Proc. R. Soc. Math. Phys. Eng. Sci. **470**, 20130819 (2014)
15. F. Frank, B. W. McArdell, C. Huggel, and A. Vieli, Nat. Haz. Earth Syst. Sci. **15**, 2569 (2015)
16. M. Mergili, M. Jaboyedoff, J. Pullarello, and S. P. Pudasaini, Nat. Haz. Earth Syst. Sci. **20**, 505 (2020)
17. J. E. Gartner, S. H. Cannon, and P. M. Santi, Eng. Geol. **176**, 45 (2014)