Machine Learning for Improved Post-fire Debris Flow Likelihood Prediction

Daniel Roten
San Diego Supercomputer Center
University of California San Diego
La Jolla, CA, United States
d1roten@ucsd.edu

Jessica Block
San Diego Supercomputer Center
University of California San Diego
La Jolla, CA, United States
jblock@ucsd.edu

Daniel Crawl
San Diego Supercomputer Center
University of California San Diego
La Jolla, CA, United States
lcrawl@ucsd.edu

Jenny Lee
San Diego Supercomputer Center
University of California San Diego
La Jolla, CA, United States
jjl053@ucsd.edu

Ilkay Altintas
San Diego Supercomputer Center
University of California San Diego
La Jolla, CA, United States
ialtintas@ucsd.edu

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020574

***Abstract*—Timely prediction of debris flow probabilities in areas impacted by wildfires is crucial to mitigate public exposure to this hazard during post-fire rainstorms. This paper presents a machine learning approach to amend an existing dataset of post-fire debris flow events with additional features reflecting existing vegetation type and geology, and train traditional and deep learning methods on a randomly selected subset of the data. The developed methods achieve AUC (area under the receiver operational characteristic curve) values of 0.93 (random forest) and 0.92 (neural network) on the test set, representing a significant improvement over a logistic regression model currently used (AUC 0.79). The paper also overviews a distributed, Kubernetes-based big data processing pipeline to efficiently retrieve features in areas impacted by new fires, and deploy the methods for real-time prediction of debris flow hazards.**

***Index Terms*—machine learning, parallel big data processing, debris flow, wildfire**

I. Introduction

Climate change and the expansion of settlements into wildfire-prone areas have resulted in droughts and heatwaves, increasing the severity of wildfires and their impact on human life, properties and ecosystems. The destruction of vegetation during such severe wildland fires can lead to debris flows in steep terrain if extreme rainfall occurs within a couple of years of the fire. Climate models predict a significant increase in the number of extreme fire weather events followed by extreme rainfall events by the turn of the century [1]. In order to enable decision makers, emergency managers and the public to properly prepare and respond to this kind of cascading hazard, it is imperative to develop reliable models for the prediction of post-fire debris flow likelihood.

One such model was developed by Staley et al. [2] based on terrain, burn severity and precipitation variables collected from the Transverse and Peninsular ranges in Southern California. The Staley et al. [2] model (referenced as ST16 through the remainder of this text) is currently being used by the U.S. Geological Survey (USGS) for emergency assessment of post-fire debris-flow hazards [3]. The ST16 model correctly predicted a high debris flow likelihood (65% for a peak 15 minute intensity of 24 mm/h) in the Montecito area (Southern California) after the 2017 Thomas Fire and before the arrival of a winter storm [4]. On January 9, 2018, heavy rainfalls on the burn area above Montecito triggered several debris flows, which resulted in the loss of 408 structures and 23 lives.

However, application of the ST16 model to slopes in the Klamath Mountains (northern California) overestimated debris flow likelihoods following the 2018 Carr and Delta Fires [5], likely due to differences in vegetation, rainfall, and underlying hydrogeomorphic processes between the Klamath Mountains and the Transverse and Peninsular ranges where the data used to train the ST16 model were collected.

In this work, we explore the use of different machine learning (ML) techniques to more accurately predict post-wildfire debris flow likelihoods throughout the U.S. We expand an existing debris flow dataset with new features with the aim of capturing differences between debris flow sites from different regions. These features are extracted from large high-resolution elevations, fuel and geological raster and vector data for hundreds of sites. A distributed big data ingestion pipeline is developed, which efficiently computes these features from past and new wildfire-affected areas for rapid emergency assessment of debris-flow likelihood.

Section II discusses the ST16 dataset and debris flow likelihood model. Section II also analyzes the additional features we added to the dataset. Section III presents the distributed big data processing pipeline developed for the extraction of new features. The performance of different ML models is presented in Section IV. Finally, Section VI shows how predictions from the presented preferred debris flow likelihood model compare

This work was supported in part by NSF awards CNS-1730158, ACI-1540112, ACI-1541349, OAC-1826967, OAC-2112167, CNS-2120019, the University of California Office of the President, and UCSD CalIT2/QI for PRP.

Abbreviation	Unit	Description	Importance (rank)
Rainstorm intensity features:			
i15*	mm/h	Peak 15-minute storm intensity	0.21 (1)
accum§	mm	Total storm accumulation	0.16 (2)
duration§	h	Total storm duration	0.10 (4)
Base site features:			
PropHM23*	-	Proportion of basin w/ slope $> 23°$ and dNBR > 440	0.10 (5)
Area§	km^2	Area of watershed (basin, catchment)	0.07 (7)
KF*	-	Soil erodibility factor K_{fact} (from STATGO database)[6]	0.07 (8)
dNBR1000*	-	Differential normalized burn ratio, divided by 1000	0.06 (9)
Additional site features:			
FFL+	t/ac	Mean fine fuel load in basin	0.11 (3)
SuscFrac+	-	Proportion of basin covered by debris-flow susceptible vegetation classes	0.08 (6)
SedUn+	-	Proportion of basin underlain by sedimentary rocks or unconsolidated deposits	0.03 (10)
Label:			
response*	-	0: no debris flow, 1: debris flow	

TABLE I: Model features and label used for debris flow prediction in this study. *Features/label used in ST16 model and provided by ST16 dataset, §features provided by ST16 dataset but not used in ST16 model, +features added in this study. The features relative importance (and rank) apply to a random forest classifier trained with the additional features.

against the ST16 model.

II. ST16 Debris Flow Data and Likelihood Model

The ST16 dataset contains 1,243 complete records from 716 debris flow sites distributed throughout the Intermountain West of the Western United States and Southern California. The empirical data combines geospatial features of the basin (also called watershed or catchment; used interchangeably in this paper), which is defined as the region from which precipitation draining into the debris flow site is collected. These features (Table I) reflect basin steepness (PropHM23), burn severity (dnbr1000) and soil properties (KF). The data also reflect features of the rainstorm during which the event (debris flow or no debris flow) was recorded, such as peak intensity in a 15 minute interval (i15), storm duration and total accumulation.

The ST16 model predicts the likelihood of a debris flow P_{df} using a logistic regression from the multiplicative combination of four of these features:

$$P_{\text{df}} = S\big(-3.63 + 0.41 \cdot \text{PropHM23} \cdot \text{i15} + 0.67 \cdot \text{dNBR1000} \cdot \text{i15} + 0.7 \cdot \text{KF} \cdot \text{i15}\big), \quad (1)$$

where $S(x) = 1/1+e^x$ is the sigmoid function. These engineered features were chosen by ST16 to force zero debris flow likelihood probability in the absence of rainfall.

Although the ST16 dataset contains data from 7 different states (NM, CA, UT, CO, MT, AZ, and ID), the model was trained only using Southern California data, with data from remaining regions used for testing. This train-test split was chosen by ST16 because the Southern California data were considered to be of the highest quality and consistency, as they originated from a single agency (the USGS); while the remaining data were collected by different sources and agencies. ST16 also found the best performance using this train-test split. However, the Southern California data span a more limited range of rainfall intensities (Fig. 1(a)) than the remaining sites (Fig. 1(b)) and show generally a lower rainfall intensity threshold than the remaining sites. This discrepancy might partially explain the overestimation of debris flow likelihoods in Northern California by Staley et al.'s model.

ST16 measured the model's performance using the threat score (a.k.a., Jaccard index in binary classification), which is the ratio between true positives (TP) and the sum of true positives, false positives (FP) and false negatives (FN):

$$T = \frac{\text{TP}}{\text{FN} + \text{FP} + \text{TP}} \quad (2)$$

This measure is preferable over the accuracy due to the biased composition of the dataset, which contains 334 debris flow events and 1,216 non debris flow events. ST16 reported a threat score of 0.42 on the training (Southern California) and 0.39 on the testing (remaining regions) sites.

Definition of Additional Features

We explore the use of new basin features in order to capture differences in debris flow response between different regions. These features target the vegetation covering the watershed and the geology of the underlying rock.

A. Fraction of Susceptible Vegetation Types

The type of vegetation affects parameters such as soil retention, burn intensity and therefore post-fire debris flow potential. For example, burnt grasslands are typically characterized by high dNBR1000 (Table I) despite low burn severity while burnt forests with largely intact canopies but significant surface fuel consumption may exhibit a low dNBR1000 despite high burn severity [e.g. 7]. In order to allow ML methods to learn these dependencies, we added a feature reflecting vegetation type to the data. We represent basic vegetation type using the 40 Scott and Burgan Fire Behavior Fuel Model (FBFM40) category [8], which is provided on a 30 meter raster by LANDFIRE [9, 10]. Figure 2(a) compares non debris-flow and

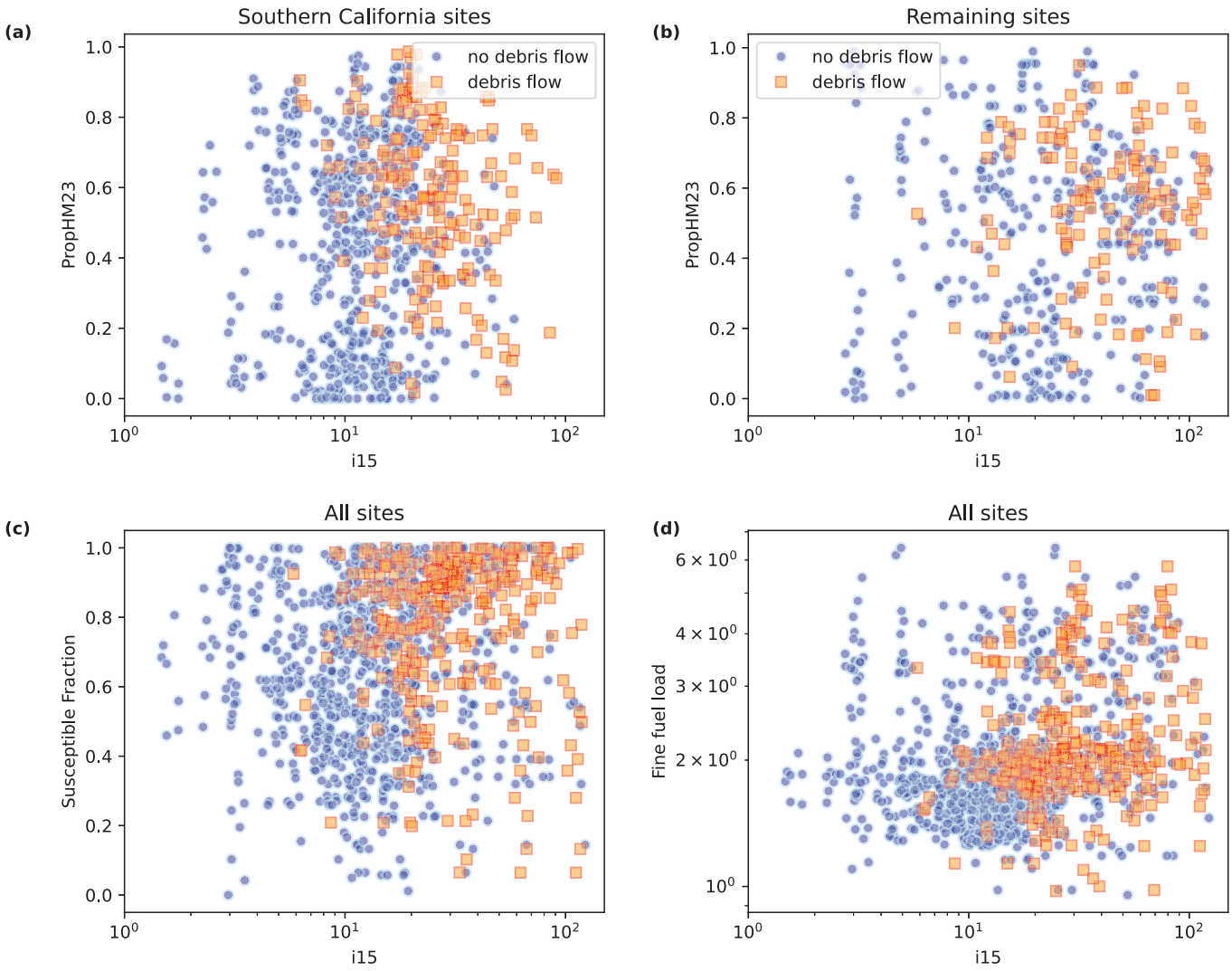


Fig. 1: Location of no debris flow and debris flow observations in feature space. Peak 15 minute rainfall intensity i15 vs. steepness PropHM23 for *(a)* Southern CA sites; *(b)* Remaining sites; *(c)* i15 vs. fraction of susceptible vegetation types (SuscFrac) for all sites; and *(d)* i15 vs. fine fuel load derived from FBFM40 model for all sites.

debris-flow occurrences in the dataset by dominant FBFM40 fuel model category.

Events in grasslands (GR) categories and grasslands and shrubs (GS) categories make up most of the observations (Figure 2a), but debris flows are more common in grasslands and shrubs (30%) than in grasslands (13%). To limit the total number of features in the ML model, we combined the vegetation categories GS, SH (shrubs), TL (timber litter) and TU (timber understory) from the FBFM40 fuel model into a new feature, called the fraction of susceptible vegetation types (SuscFrac, Table I). Watersheds with a higher SuscFrac value appear to exhibit debris flows more often than watersheds with a low fraction (Figure 1(c)).

B. Fine Fuel Load

The FBFM40 fuel model subdivides each fuel type into different subcategories with different fuel characteristics, specified as fine fuel load (FFL) per unit area, surface area volume, packing ratio and extinction moisture content. For example, grasslands range from sparse grasses in category GR1 with a FFL of 0.4 t/ac (tons per acre) to dense, tall grasses in category GR9 with a FFL of 10 t/ac. We found that the mean FFL within the watershed provides the greatest performance improvement (compared to the other fuel metrics in FBFM40) in the ML models and added it as a new feature (Table I). Higher mean FFLs tend to be associated with more frequent debris flow occurrences (Figure 1(d)). This observation could be related to the higher burn severity caused by the added fuel, which leads to more debris and further reduces the soil's capability to infiltrate precipitation.

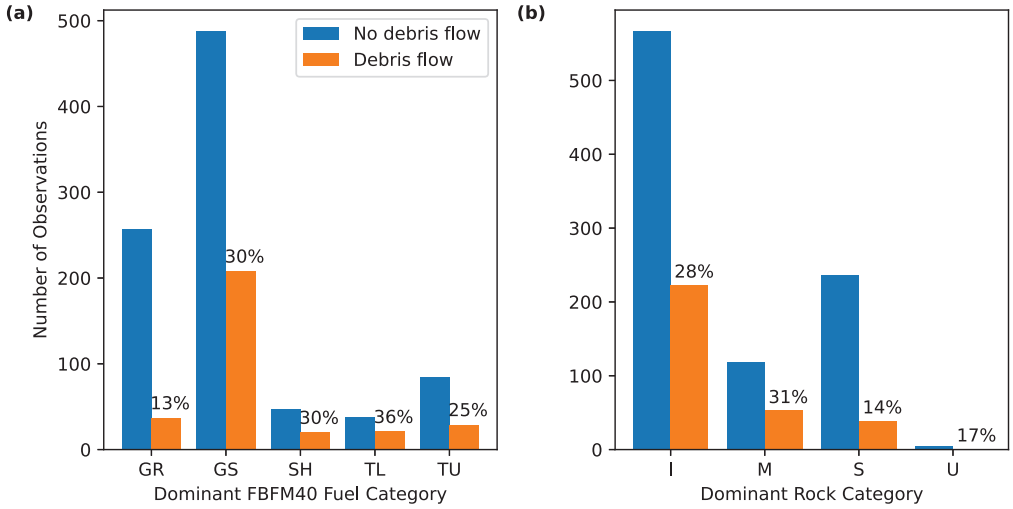


Fig. 2: Contribution of no-debris flow and debris flow events by *(a)* FBFM40 fuel category and *(b)* Rock category. GR = grassland, GS = grassland-shrubs, SH = shrubs, TL = timber-litter, TU = timber-understory. I = igneous, M = metamorphic, S = sedimentary, U = unconsolidated.

C. Fraction of Sedimentary and Unconsolidated Rocks

Debris flow likelihood also depends on the type of underlying rock. We queried the digital version of the geological map of each state [11] to calculate the fraction of each rock class (igneous, metamorphic, sedimentary and unconsolidated) in the watershed. In the ST16 dataset, watersheds dominated by igneous and metamorphic rocks are subject to debris flows more frequently than watersheds dominated by sedimentary rocks and unconsolidated deposits (Figure 2(b)). We defined the fraction of covered by sedimentary rock and unconsolidated deposits as a new feature (SedUn, Table I).

III. Debris Flow Big Data Processing Pipeline

Operational debris flow forecasting essentially represents a big data application, as it depends on features derived from a large volume and variety of data, as well as features that may change quickly during emergencies (velocity). The USGS digital elevation model (DEM) used for watershed calculation, for example (described below), contains several Tb of data (although only a small fraction of that data covers an individual watershed). Feature extraction also processes data of multiple types. Geological maps, for instance, are provided as structured data in shapefile format, while fuel models or DEMs are stored in raster format representing unstructured content. The data originate from different agencies and projects (e.g., USGS for DEM or geological maps, U.S. Department of Agriculture, Forest Service for fuel model). The final debris flow training set with the additional features represents a compilation of data collected from a wide range of different sensor types, including

- aerial or satellite IR images for the calculation of burn ratios (dNBR1000) and definition of fire perimeters,
- satellite visible spectrum images (Landsat) for the fuel (LandFire) and soil models (KF),
- satellite Lidar data for the DEM,
- rain gauges for precipitation monitoring [ST16],
- video cameras for debris flow monitoring [12], and
- field observations for geologic maps or debris flow response.

As precipitation intensity forecasts evolve during post-fire storm situations, debris flow likelihood estimates also need to be updated to these frequently changing features.

A distributed data processing pipeline was developed to rapidly provide features for training and deploying the ML models described below (Figure 3). Feature extraction operations for different watersheds were distributed on the Nautilus Kubernetes cluster of the Pacific Research Platform (PRP) [13, 14] using Dask [15, 16] bags for Python. The Helm package manager for Kubernetes [17] was used to install Dask on Nautilus. Watersheds pertaining to each debris flow location were computed using the PySheds [18] library and a DEM from the USGS 3DEP program provided in $1/3$ arc second resolution [19]. Raster data, such as the DEM and fuel model [9] were processed using the Xarray library [20, 21] with the rioxarray extension [22]. Vector data such as watershed polygons and the digital representation of the geological map [11] were processed using Geopandas dataframes [23]. Key operations carried out by the data processing pipeline include (Fig. 3):

- For each watershed area:
  - Retrieve DEM for site surroundings
  - Extract watershed polygon for site
  - Retrieve fuel model raster for region
  - Re-project DEM to match fuel model reference system and resolution
  - Perform spatial join between watershed area and fuel model

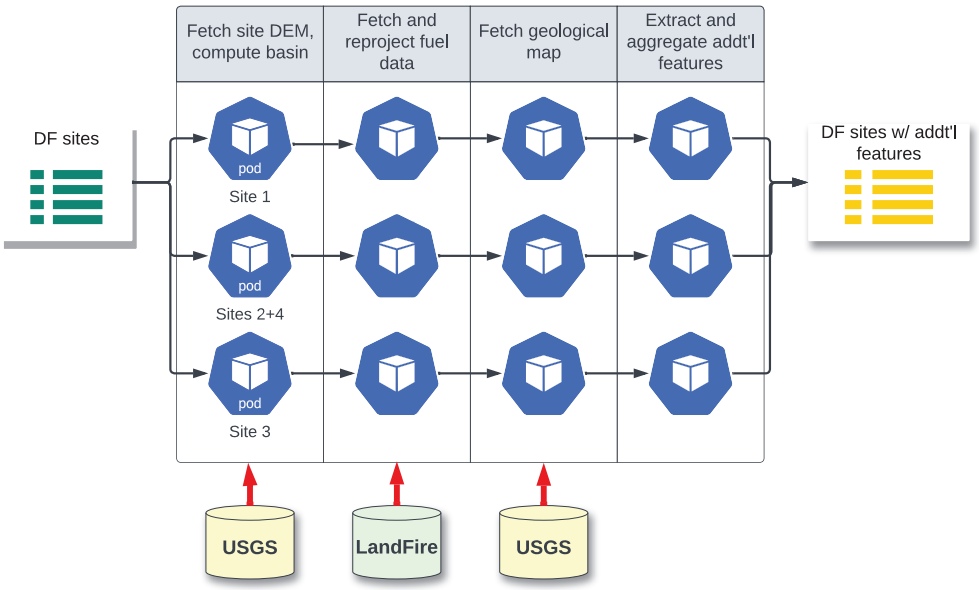


Fig. 3: Simplified schematic of data processing pipeline for computation of basins and extraction of additional site features. DF = debris flow.

  - Aggregate fuel parameters (SuscFrac, FFL) for watershed
  - Retrieve geological map for site's state
  - Carry out spatial join between watershed and geological map
  - Aggregate by geological classification to compute SedUn.
- Collect features from each watershed
- Join with debris flow event data and write to disk.

The feature extraction pipeline uses a data parallel programming model. Only site metadata (including location) is defined on the scheduler node and distributed to the worker nodes, with each worker running in a separate pod. Variables with large memory footprints, such as the DEM, the fuel model and the geological map, are stored in each worker's local memory. A persistent volume claim (PVC) on the Kubernetes cluster is mounted by all pods and used to stage geological maps, which are downloaded only once for each required state. The Landfire fuel model was converted to a cloud-optimized GeoTiff (COG) format to reduce memory usage and also staged on the PVC. The DEM data is already provided in COG format, and the relevant area is directly retrieved by each worker from the USGS servers using HTTP range requests and the rioxarray interface [22]. The data extraction and feature aggregation for the site's watershed is carried out individually on each worker node (Fig. 3), and only the condensed site features (Table I) and the watershed polygon are collected by the scheduler and stored to disk. Using 10 pods on Nautilus, the feature extraction completes in $\sim$9 minutes for the 716 debris flow sites in the ST16 dataset. However, debris flow prediction during active, large scale wildfires may be required for several thousand potential watersheds, which would necessitate the use of more pods.

IV. ML Models and Performance

We evaluated the performance of different ML models by splitting the dataset into training and testing sites.

A. Train-Test Split

As we want the model to generalize well to sites outside of Southern California, sites were randomly assigned as training or testing sites regardless of their location. Because more than one observation was made at most debris flow sites, the split between training and testing data was done by site identifier, rather than by observation. This avoids splitting observations made at the same site between the training and testing set, and tests the realistic situation of predicting debris flows at sites not used for training. The training set contains 982 observations collected at 522 (80%) sites, while the test set contains 261 observations collected at 131 (20%) sites.

B. Addition of Random Noise to Rainstorm Features

Precipitation observations in the dataset were collected from rain gages up to 4 km away from the watersheds [ST16], and only 123 unique precipitation values are present in the dataset. The assumption that the data are independent and identically distributed may thus not hold for this dataset. We found that random forests in particular tend to overfit to small-scale dependencies of debris flow likelihoods on precipitation data during training. Nearby watersheds sharing the exact same i15 value may exhibit a similar debris flow response, and the ML algorithm effectively learns to assign a debris flow site to a group of similar sites based on the unique i15 record. We

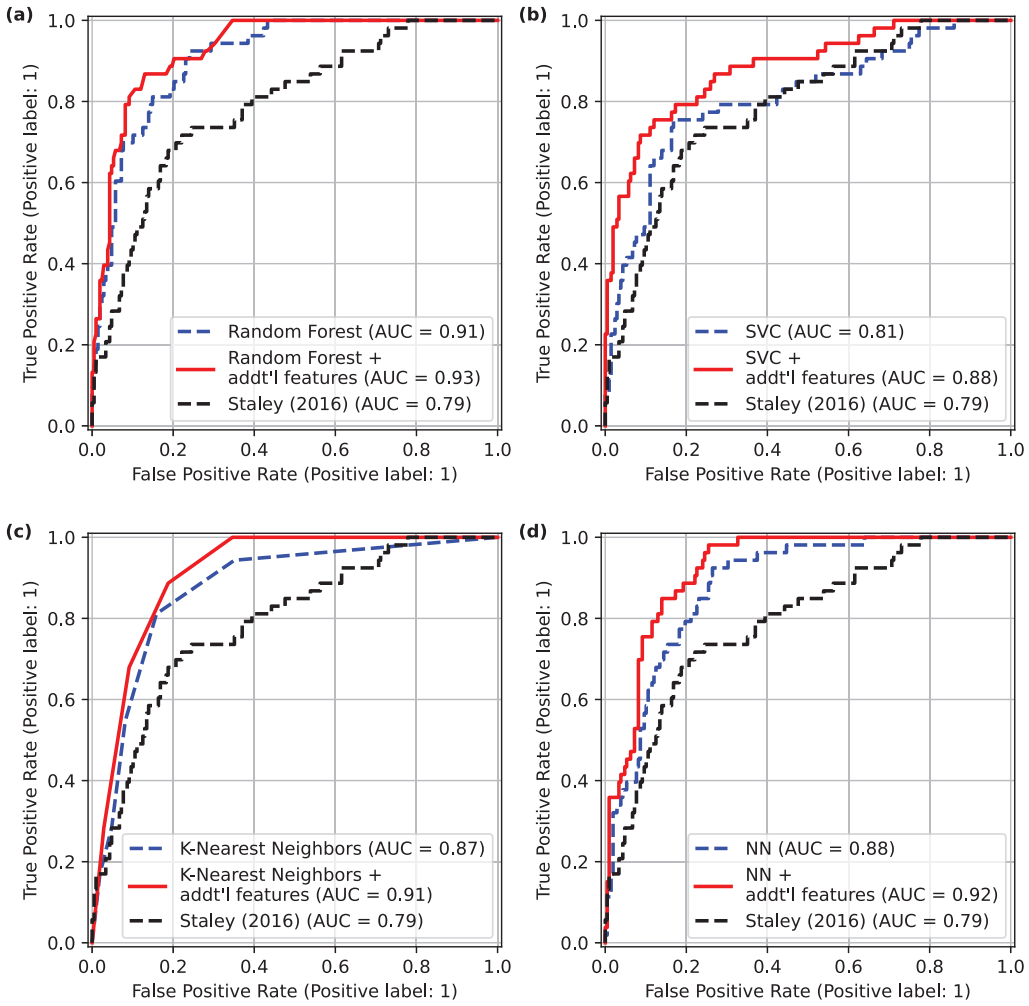


Fig. 4: Receiver operating characteristics (ROC) curves obtained on the test sites for *(a)* Random forest, *(b)* Support vector classifier, *(c)* K-nearest neighbors, and *(d)* neural network. ROCs are shown using features from ST16 dataset and using three additional features (SuscFrac, FFL and SedUn). The ROC curve of the ST16 model is shown as baseline.

therefore added random noise to the precipitation features (i15, duration, total accumulation) before the test-train split (Fig. 1). The amplitude of the noise was defined to range between -10% and 10% of the recorded value.

C. Performance Evaluation

We trained different ML methods on the training set, and recorded the threat score and area under the receiver operating characteristics (ROC) curve during training and testing (Table II). Performance was evaluated using the base features (Table I) and using the base and additional features (SuscFrac, FFL, and SedUn) with the same train-test split.

Traditional ML models including logistic regression, naive Bayes, support vector classifiers (SVC), K-nearest neighbors (KNN) and random forests were implemented using the Scikit-learn package [24]. Hyperparameters such as the regularization in the SVC and the maximum tree depth in the random forest were calibrated to optimize threat scores on the test set. For the SVC, KNN and the neural network, features were standardized by removing the mean and scaling to unit variance.

Method	base features		w/ addt'l features	
	training	testing	training	testing
ST16	0.39 (0.80)	0.37 (0.79)		
Logistic regr.	0.35 (0.83)	0.32 (0.86)	0.35 (0.84)	0.36 (0.87)
Naive Bayes	0.24 (0.79)	0.27 (0.82)	0.27 (0.79)	0.31 (0.84)
SVC	0.70 (0.97)	0.36 (0.81)	0.75 (0.98)	0.42 (0.88)
KNN	0.59 (0.94)	0.41 (0.87)	0.61 (0.95)	0.50 (0.91)
Random Forest	0.97 (1.00)	0.45 (0.91)	1.00 (1.00)	0.53 (0.93)
Neural Network	0.59 (0.93)	0.37 (0.88)	0.75 (0.98)	0.54 (0.92)

TABLE II: Training and test threat scores (and area under the curve) of different ML methods using base features and with additional features SuscFrac, FFL and SedUn. SVC = support vector classifier, KNN = K-Nearest Neighbors.

The neural network (NN) was implemented in Python using the Keras [25] interface to the Tensorflow [26] backend. We used a fully connected architecture, with an input node for

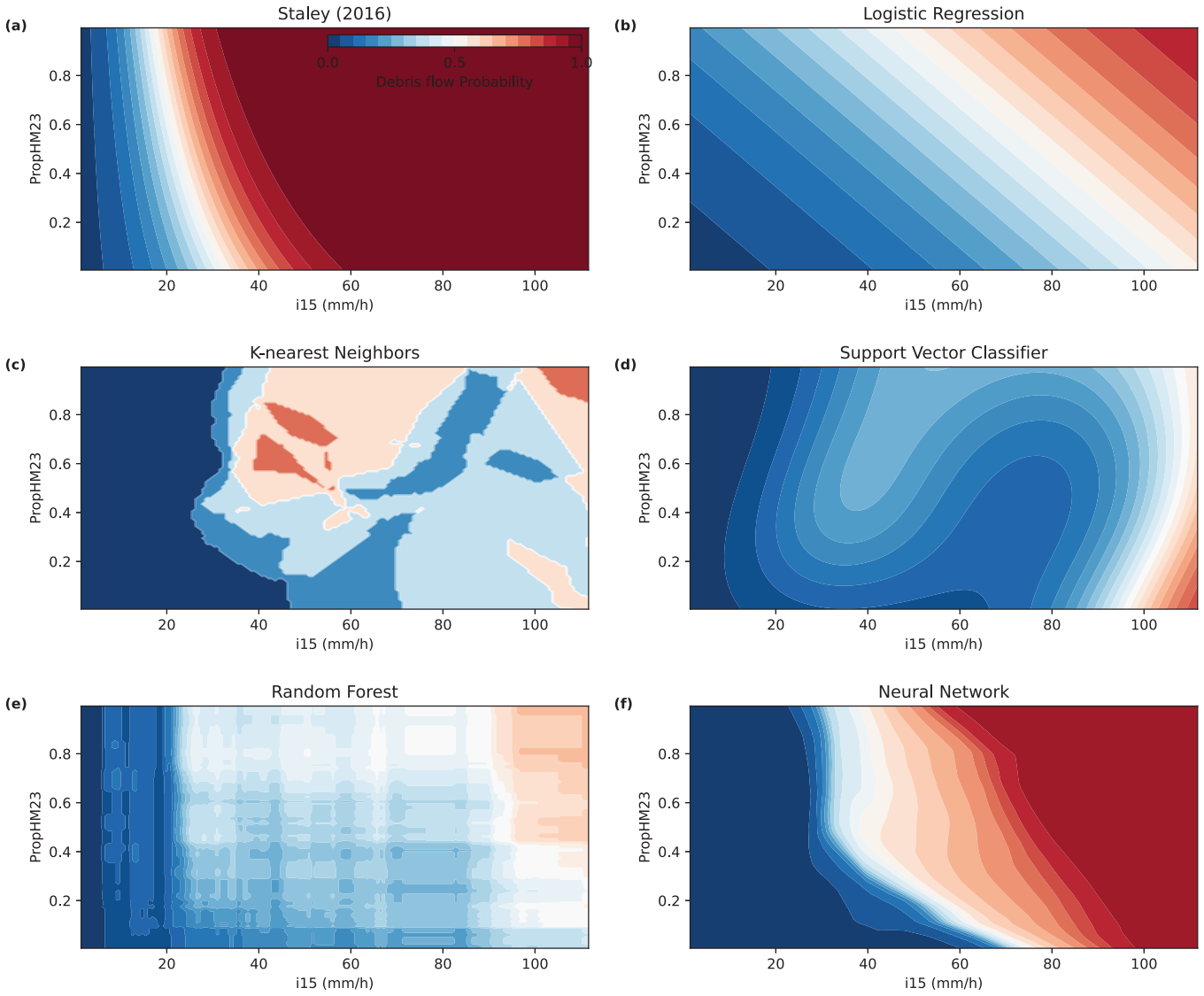


Fig. 5: Debris flow probability as function of i15 (mm/h) and PropHM23 predicted using *(a)* ST16, *(b)* logistic regression, *(c)* KNN, *(d)* SVC, *(e)* random forest, and *(f)* neural network. Other features were set to constant values except for the total storm accumulation. See text for more explanations.

each feature, several hidden layers and a single output layer. Drop-out regularization was applied to reduce overfitting to the training data. The number of hidden layers, layer size and drop out probabilities were established by trial and error to optimize performance. A design with 7 hidden layers and up to 92 nodes per layer was used for the NN with the base features. The NN trained with the additional features uses 11 hidden layers with up to 392 nodes per layer. Hidden layer weights were initialized using the Xavier (Glorot) method to minimize problems with vanishing or exploding gradients [27]. Drop-out probabilities were calibrated to values between 0.2 and 0.3. The activation functions alternate between tanh and rectified linear units in the hidden layers [28, 29]; the output layer uses a sigmoid function to predict the probability of a debris flow.

The ST16 model achieves a threat score of 0.37 on the test set (Table II), slightly lower than the value of 0.39 reported by the creators on the observations outside of Southern California. Without the additional features, the threat scores of the SVC and the NN are on par with ST16, while the logistic regression and Naive Bayes classifier score lower at 0.32 and 0.27, respectively. The random forest achieves 0.45. The additional features improve threat scores for all the tested ML methods, with the KNN, random forest and NN scoring higher than 0.50 (Table II).

All the tested ML methods achieve AUCs (area under the receiver operational characteristic curve) above 0.80 even without additional parameters, while ST16 achieves 0.79 (Fig-

ure 4, Table II). With the additional features, we recorded an AUC of 0.92 with the NN and an AUC of 0.93 with the random forest.

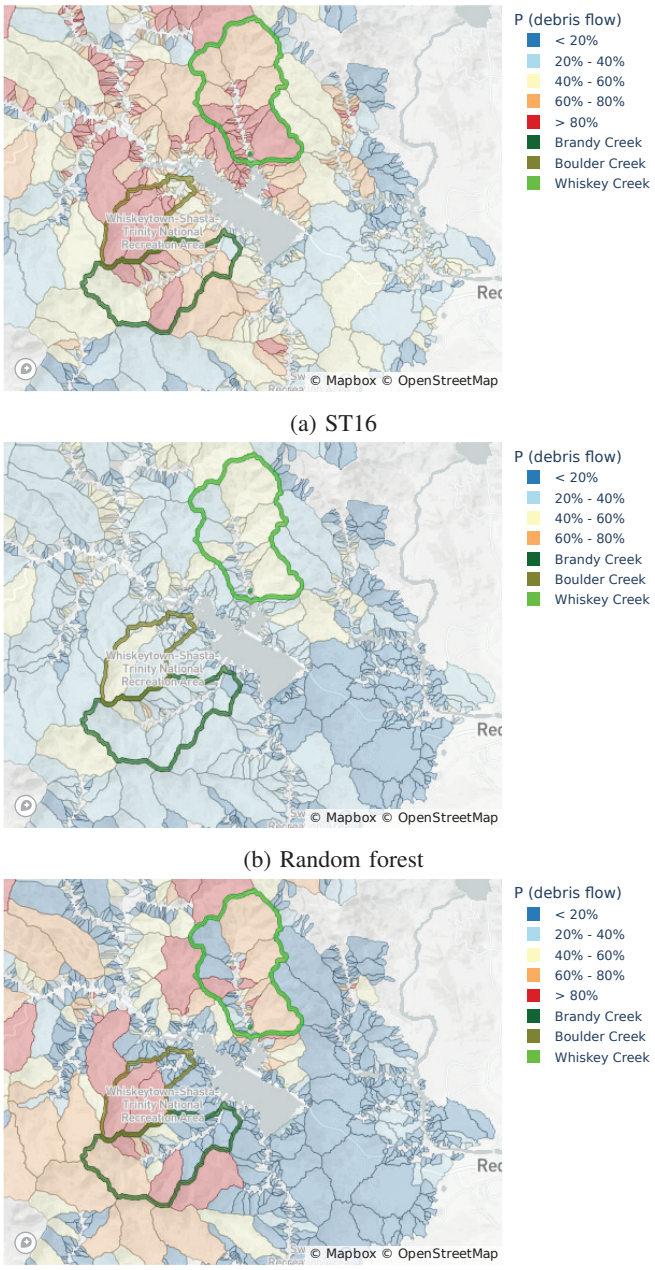
These improvements are significant for operational debris flow forecasting, where the true positive rate is to be maximized and the false positive rate (FPR) is to be minimized, as frequent false alarms may desensitize stakeholders to alerts of a real threat. Catching 80% of debris flows would result in an FPR of 8% using a random forest trained with the additional features, but 15% using the random forest trained on just base features (Figure 4(a)). Using the neural network, FPRs would be 13% and 22%; respectively (Figure 4(d)). With the original ST16 model, setting the threshold for a true positive rate to 80% would incur an FPR of 38% (Fig. 4).

The computational resources required by the additional features during training and testing are not relevant. The traditional machine learning models (e.g., SVC, random forest) take less than one second for training. Due to the larger number of hidden layers, the NN with the additional features takes longer ($\sim$3 minutes) to train than the NN without those features ($<$1 minute); however, predicting the debris flow likelihood for the Carr fire sites took less than a second for all models and all features.

V. Interpretation of ML Results

To explore the variability of the predicted debris flow likelihoods with changing features, we calculated probabilities from each classifier through the feature space spawned by the most important rainfall feature, i15 (Table I), and the most important non-precipitation base feature, PropHM23 (Figure 5). The total storm accumulation was set to 0.4$\times$i15 following the general trend in the data. The remaining features were set to constant values, with the storm duration set to 12 hours, the watershed area to 1 km^2, dnbr1000 = 0.41, kf = 0.25, SedUn = 0, SuscFrac = 0.98 and FFL = 2.0; these values are close to the median values of the watersheds analyzed in Section VI.

The ST16 (Figure 5(a)) model predicts a debris flow probability of $P_{\text{df}} = 1$ for about two thirds of the feature space, which does not fully reflect the observations especially at sites outside Southern California (Figure 1(b)). The logistic regression (Figure 5(b)) predicts a more gradual increase in P_{df}, but gives $P_{\text{df}} > 0$ for zero precipitation, a problem already noted by ST16. The remaining shallow and deep ML methods (Figure 5(c-f)) return $P_{\text{df}} \approx 0$ for i15 ≈ 0. Predictions by the K-nearest neighbors (Figure 5(c)) do not monotonically increase with increasing precipitation. Although this behavior reflects patterns in the underlying data, such predictions can be potentially problematic in real-world scenarios. The random forest does also not strictly exhibit increasing P_{df} with increasing i15, but this effect is strongly attenuated due to the randomization inherent to this method. The P_{df} distribution of the NN resembles the distribution from the logistic regression, but predicts a more complex dependency on the two varied features. The distribution of P_f as a function of i15 and PropHM23 obtained by the SVC (Figure 5(d)) seems at odds with the distribution suggested by the other methods. For these



(a) ST16

(b) Random forest

(c) Neural network

Fig. 6: Debris flow likelihoods in Shasta County following the 2018 Carr Fire during an assumed rainstorm with i15 = 24 mm/h, duration = 12 h and a total accumulation of 20 mm. Green colors show the watersheds surrounding the three creeks monitored by East et al. [30].

reasons, we selected the random forest and neural network as our preferred methods for operational predictions of post-fire debris-flow likelihood.

VI. Debris Flow Likelihood Post 2018 Carr Fire

The training and test sets were merged into a single dataset, and used to re-train the different ML models with the hyper-parameters selected during testing.

We then calculated debris flow likelihoods following the 2018 Carr Fire in the Whiskeytown Lake (Shasta County) area to illustrate the sensitivity of the predictions to the ML method. East et al. [30] monitored sediment transport from three basins draining into Whiskeytown Lake during the first two years following the fire and observed no post-fire debris flows. However, multiple rainstorms with peak 15 minute intensities above 24 mm/h were observed by rainbow gauges installed within the basins, which would have brought the debris flow likelihoods according to Staley et al.'s model above 60% in almost all of the basins surrounding the three monitored creeks (Figure 6a).

We retrieved watershed extents and base site features (Table I) from the USGS emergency debris flow assessment website [3]. Additional features (Table I) were calculated using a slightly modified version of the workflow developed for the preparation of training and testing data (Section III).

For the storm features, we assumed a peak 15m precipitation intensity of 24 mm/h, a storm duration of 24 hours and total accumulation of 20 mm. Figure 6 shows the debris flow likelihoods predicted by the different ML methods.

The random forest predicts $P_{\mathrm{df}} < 60\%$ in all basins except a small one within Brandy Creek (Figure 6b), with most watersheds having P_{df} between 20 and 40%. The debris flow likelihoods predicted by the neural network (Fig. 6c) are between those from the random forest and the ST16. Both the random forest and the NN predict lower likelihoods than ST16, in particular in the less steep areas downstream from the lake. These results are generally consistent with observations of the P_{df} within the feature space (Fig. 5), and confirm that the ST16 model tends to overpredict debris flow likelihoods [5]. However, the large number of basins with elevated P_{df} ($> 40\%$) would make it unlikely that no post-fire debris flows were triggered within two years [30], in particular considering that some rain gauges recorded i15 values well above 24 mm/h.

VII. Interactive Maps of Debris Flow Likelihood

The distributed feature extraction pipeline allows for real-time prediction of debris flow likelihoods in fire affected regions. We are currently developing an interactive web application that displays color-coded maps of debris flow likelihoods for changing assumptions (Figure 7). Users are able to explore the sensitivity of P_{df} to storm (precipitation and duration) parameters and the choice of ML method. Future versions will couple the interactive map with live weather forecasts, thereby giving stakeholders the opportunity to react to rapidly evolving storm situations in fire-ravaged regions.

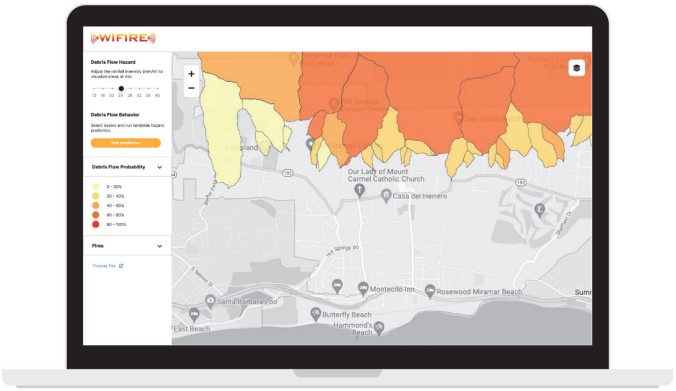


Fig. 7: Example of interactive web application, showing map of post-fire debris flow likelihood following the 2017 Thomas Fire.

VIII. Conclusions

We have developed a distributed big data ingestion pipeline for the extraction of additional basin features in a published dataset of post-fire debris flow events. The additional features include the fraction of susceptible vegetation type in the basin (SuscFrac), the mean fine fraction of the fuels covering the basin (FFL), and the fraction of sedimentary and unconsolidated rocks underlying the basin (SedUN). Adding these extra features significantly improves the threat score and area under the ROC curve using both traditional ML methods and a fully connected NN. A random forest and NN with 11 hidden layers are our preferred models for debris flow prediction. Both methods predict significantly lower debris-flow likelihoods in basins surrounding Whiskeytown Lake following the 2018 Carr Fire, bringing the predictions closer to observations. An interactive web application is under development which will allow decision makers to consider these alternative debris-flow likelihood models when assessing the dangers posed by large precipitation events over fire-affected basins.

Acknowledgement

The data processing pipeline was deployed on the Nautilus Kubernetes Cluster of the Pacific Research Platform [13] using name-space "wifire-quicfire". The authors thank Mai Nguyen for her helpful comments, which helped to improve the manuscript. The authors are grateful to Abani Patra and Luke Mcguire for the successful collaboration that contributed to this research. The authors thank the four anonymous reviewers for their helpful suggestions.

References

[1] Danielle Touma, Samantha Stevenson, Daniel L Swain, Deepti Singh, Dmitri A Kalashnikov, and Xingying Huang. Climate change increases risk of extreme rainfall following wildfire in the western United States. *Science advances*, 8(13), 2022.

[2] Dennis M Staley, Jacquelyn A Negri, Jason W Kean, Jayme L Laber, Anne C Tillery, and Ann M Youberg. *Updated logistic regression equations for the calculation of post-fire debris-flow likelihood in the western United States*. U.S. Geological Survey, 2016.

[3] U.S. Geological Survey. Emergency assessment of post-fire debris-flow hazards. URL https://landslides.usgs.gov/hazards/postfire_debrisflow/.

[4] Jason W Kean, Dennis M Staley, Jeremy T Lancaster, Francis K Rengers, Brian J Swanson, Jeffrey A Coe, JL Hernandez, AJ Sigman, Kate E Allstadt, and Donald N Lindsay. Inundation, flow dynamics, and damage in the 9 January 2018 Montecito debris-flow event, California, USA: Opportunities and challenges for post-wildfire risk assessment. *Geosphere*, 15(4):1140–1163, 2019.

[5] Donald Lindsay, David Cavagnaro, Nathan Delgado, Scott W McCoy, and Jason W Kean. Post-fire runoff response in northern California: Testing current models used to predict post-fire runoff magnitude and debris flow triggering thresholds. In *AGU Fall Meeting Abstracts*, volume 2020, pages NH006–06, 2020.

[6] Soils data for the Conterminous United States derived from the NRCS state soil geographic (STATSGO) data base. URL https://water.usgs.gov/GIS/metadata/usgswrd/XML/ussoils.xml.

[7] Michael C Stambaugh, Lyndia D Hammer, and Ralph Godfrey. Performance of burn-severity metrics and classification in oak woodlands and grasslands. *Remote Sensing*, 7(8):10501–10522, 2015.

[8] Joe H Scott and Robert E Burgan. *Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2005.

[9] Landfire. URL https://landfire.gov.

[10] Kevin C Ryan and Tonja S Opperman. LANDFIRE –a national vegetation/fuels data base for use in fuels treatment, restoration, and suppression planning. *Forest Ecology and Management*, 294:208–216, 2013.

[11] U.S. Geological Survey. Geological maps of U.S. states, . URL https://mrdata.usgs.gov/geology/state/.

[12] Dennis M Staley, Anne C Tillery, Jason W Kean, Luke A McGuire, Hannah E Pauling, Francis K Rengers, and Joel B Smith. Estimating post-fire debris-flow hazards prior to wildfire using a statistical analysis of historical distributions of fire severity from remote sensing data. *International journal of wildland fire*, 27(9):595–608, 2018.

[13] *Nautilus - Pacific Research Platform*. URL https://pacificresearchplatform.org/nautilus/.

[14] Larry Smarr, Camille Crittenden, Thomas DeFanti, John Graham, Dmitry Mishin, Richard Moore, Philip Papadopoulos, and Frank Würthwein. The Pacific Research Platform: Making high-speed networking a reality for the scientist. In *Proceedings of the Practice and Experience on Advanced Research Computing*, New York, NY, USA, 2018. URL https://doi.org/10.1145/3219104.3219108.

[15] Dask Development Team. *Dask: Library for dynamic task scheduling*, 2016. URL https://dask.org.

[16] Matthew Rocklin. Dask: Parallel computation with blocked algorithms and task scheduling. In Kathryn Huff and James Bergstra, editors, *Proceedings of the 14th Python in Science Conference*, pages 130 – 136, 2015.

[17] Helm - the package manager for Kubernetes. URL https://helm.sh/.

[18] Matt Bartos. PySheds: simple and fast watershed delineation in Python. URL https://github.com/mdbartos/pysheds.

[19] U.S. Geological Survey. *3D Elevation Program*, . URL https://www.usgs.gov/3d-elevation-program.

[20] Xarray: N-D labeled arrays and datasets in Python. URL https://xarray.dev.

[21] S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1), 2017. doi: 10.5334/jors.148. URL https://doi.org/10.5334/jors.148.

[22] rioxarray. URL https://corteva.github.io/rioxarray/stable/index.html.

[23] Geopandas 0.11.0. URL https://geopandas.org/en/stable/index.html.

[24] scikit-learn: Machine learning in Python. URL https://scikit-learn.org/stable/.

[25] Keras: the Python deep learning API. URL https://keras.io/.

[26] Tensorflow: An end-to-end open source machine learning platform. URL https://www.tensorflow.org/.

[27] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[28] Phoebe MR DeVries, T Ben Thompson, and Brendan J Meade. Enabling large-scale viscoelastic calculations via neural network acceleration. *Geophysical Research Letters*, 44(6):2662–2669, 2017.

[29] Daniel Roten and Kim B Olsen. Estimation of site amplification from geotechnical array data using neural networks. *Bulletin of the Seismological Society of America*, 111(4):1784–1794, 2021.

[30] Amy E East, Joshua B Logan, Peter Dartnell, Oren Lieber-Kotz, David B Cavagnaro, Scott W McCoy, and Donald N Lindsay. Watershed sediment yield following the 2018 Carr fire, Whiskeytown national recreation area, northern California. *Earth and Space Science*, 8(9), 2021.