Supplemental Information

Impacts from cascading multi-hazards using hypergraphs: a case study from the 2015 Gorkha earthquake in Nepal

Alexandre Dunant[1*], Tom R. Robinson[2], Alexander L. Densmore[1], Nick J. Rosser[1], Ragindra Man Rajbhandari[3], Mark Kincey[4], Sihan Li[5], Prem Raj Awasthi[3], Max Van Wyk de Vries[6,7], Ramesh Guragain[8], Erin Harvey[1] and Simon Dadson[9]

[1] Institute of Hazard, Risk, and Resilience and Department of Geography, Durham University, Durham, UK
[2] School of Earth and Environment, University of Canterbury, Christchurch, New Zealand
[3] UN Resident Coordinator's Office, Nepal
[4] School of Geography, Politics, and Sociology, Newcastle University, Newcastle, UK
[5] Department of Geography, University of Sheffield, Sheffield, UK
[6] Department of Geography, University of Cambridge, Cambridge CB2 3EL, UK
[7] Department of Earth Sciences, University of Cambridge, Cambridge CB3 0EZ, UK
[8] National Society for Earthquake Technology-Nepal (NSET), Nepal
[9] School of Geography and the Environment, University of Oxford, UK

Supplementary material 1. Slope unit explanation and parameters

To define slope units, we use the methodology of Alvioli et al. (2016, 2020) as described in detail by Kincey et al. (2022) in their derivation of slope units for Nepal. We refer the reader to Kincey et al. (2022) for a full evaluation. The methodology divides a landscape into individual units separated by drainages and divide boundaries, according to a set of user-specified parameters. We use the *r.slopeunits* package within GRASS GIS v.7.8.4, run on the AW3D digital elevation model resampled to a cell size of 10 x 10 m. We choose minimum parameter settings from the range of values recommended by Alvioli et al. (2016) to ensure that smaller slope units were retained and to match observed hillslope length scales in Nepal. The resulting slope unit map is shown in Figure S1. Parameter names and values are as follows:

- The drainage area threshold *thresh* = 5,000,000 m² identifies cells with a flow accumulation greater than the threshold, delineating drainage lines and subsequently forming catchments that are subdivided into half basins (HBs).
- The minimum surface area *areamin* = 50,000 m² defines the smallest acceptable area for a slope unit.
- The minimum circular variance of terrain aspect within a slope unit *cvmin* = 0.2 plays a crucial role in determining the uniformity of HBs concerning terrain aspect. This parameter, calculated as $1 - (|R|/N_v)$, where N_v is the number of grid cells in each half basin and $|R|$ is the magnitude of the vector resulting from the summation of all unit vectors describing the orientation of each grid cell, influences the degree of homogeneity in aspect among slope units. Small values of *cvmin* result in HBs with more uniform aspect, large values in less uniformity.

- The reduction factor *rf* = 10 governs the rate at which the drainage area threshold *thresh* decreases for subsequent iterations. A smaller value of *rf* results in a faster decrease, enabling more subdivisions and iterations, albeit at the cost of processing time.
- The cleaning parameter *cleansize* = 20,000 m² sets the threshold value for cleaning procedures, removing slope units smaller than this area. It is also used in additional cleaning steps that inspect for unrealistic elongation or aspect similarity.

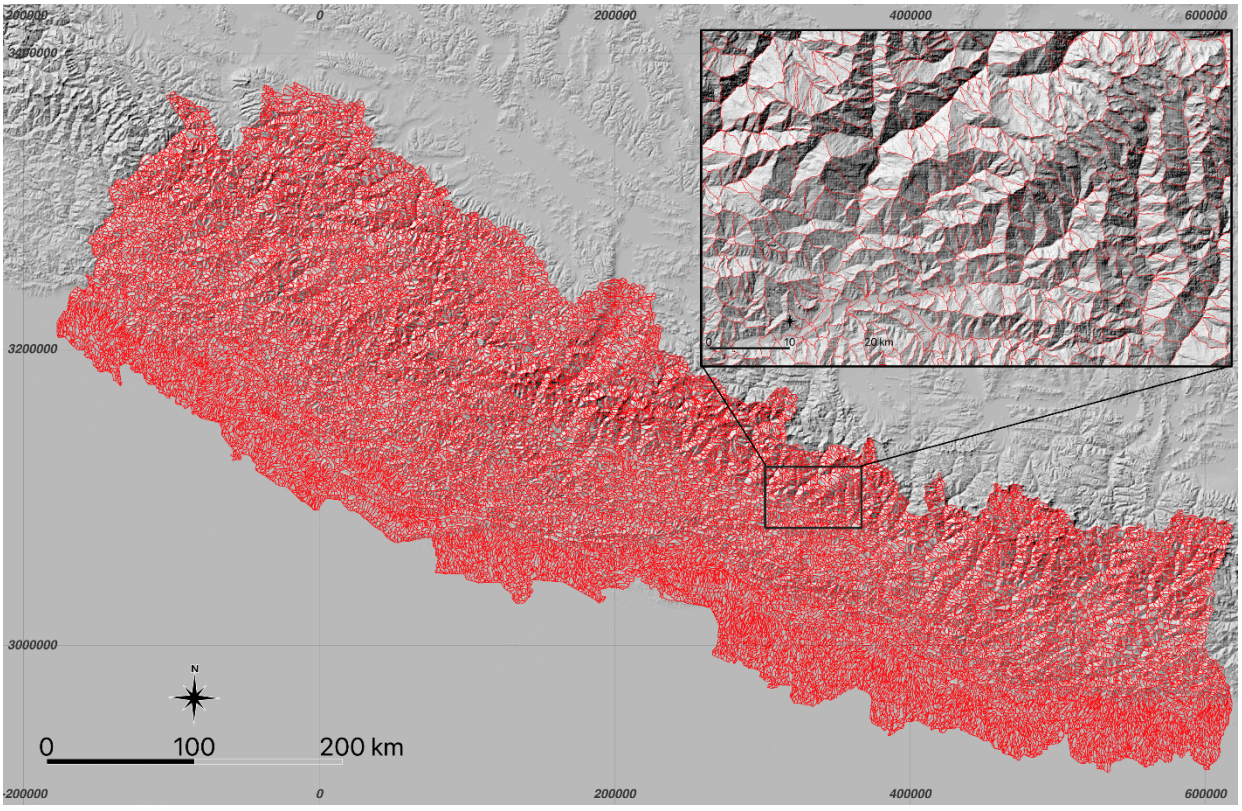


Figure S1: Map of slope units for Nepal, using the parameter values given above. The inset shows a detailed area of north-central Nepal, illustrating the detailed relationship between slope unit boundaries (red) and topography.

Supplementary material 2. Coseismic landslide susceptibility

To estimate the spatial pattern landsliding due to the 2015 Gorkha earthquake, we utilise a landslide susceptibility layer for Nepal with a cell size of 10 x 10 m that is trained on the 2015 coseismic landslide inventory. Landslide susceptibility is determined with a static susceptibility model that depends upon seven topographic factors: elevation, hillslope aspect, distance to rivers, plan-view curvature, regional relief, local hillslope gradient over a 3 x 3 cell window, and a terrain ruggedness index. All factors are derived from the AW3D digital elevation model resampled to a cell size of 10 x 10 m. *Distance to rivers* has been calculated as the straight-line (Euclidean) distance from each pixel to the nearest river and river confluence, respectively. Rivers are defined using a flow accumulation tool that identifies rivers based on an upstream area threshold of 5km^2.

Regional relief represents the standard deviation in elevations in a window surrounding each pixel using the Focal Statistics tool in ArcGIS with a 1kmx1km square window.

Terrain ruggedness index is the average difference between the elevation of the central pixel and each of the 8 adjacent pixels which is calculated using the DEM and Focal Statistics tool in ArcGIS.

We generate the susceptibility model using a gradient boosting machine learning approach, XGBoost (Chen and Guestrin., 2016), implemented in Python. The model is trained with the coseismic landslide inventory (epoch 4) of Kincey et al. (2021) which covered the 14 worst affected districts during the 2015 Gorkha earthquake; we randomly selected 20% of grid cells for model validation and used the remaining 80% of cells for model training. This yields an area under the receiver operating characteristic (ROC) curve of 0.86 (Fig. S2).

The relative importance of the features in our XGBoost-based susceptibility model was assessed through the 'feature_importances_' class from the scikit-learn Python library. This measurement is derived by evaluating how significantly each feature decreases the model's loss function. Specifically, the importance score for each feature is calculated by counting how frequently a feature is used to split the data across all decision trees, weighting these counts by the improvement (gain) brought about by each split, and normalizing these values by the total gains across the model.

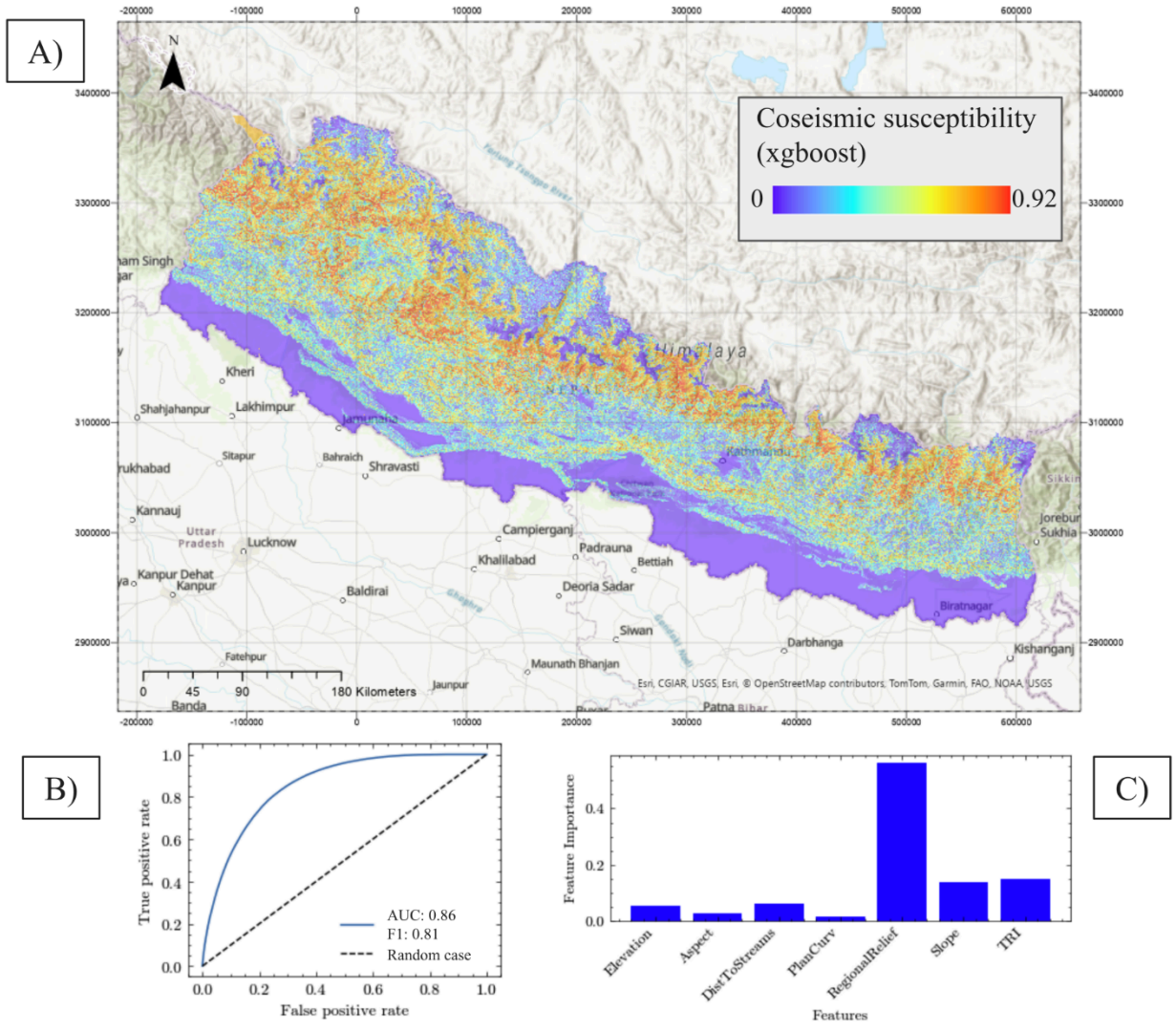


Figure S2: A, map of susceptibility to coseismic landslides derived from the static seven-parameter susceptibility model and the XGBoost algorithm. Susceptibility values range from 0 to 1. B, receiver operating characteristic curve derived for a random selection of 20% of grid cells. Values show area under the curve (AUC) and maximum F1 score. Dashed line shows

the random case (AUC = 0.5). C, relative feature importance of the seven topographic factors derived from the XGBoost algorithm. DistToStreams, distance to the nearest river; PlanCurv, plan-view curvature; TRI, terrain ruggedness index.

Supplemental material 3. Relationship between PGA and landslide occurrence

A logistic regression algorithm is used to investigate the relationship between peak ground acceleration (PGA) values and the occurrence of landslides. Initially, the script creates a feature matrix `X` containing PGA values and a target variable `y` that encodes whether a landslide occurred (1) or not (0) based on the presence of landslide events. These variables are derived from a geospatial dataset, where the PGA values are resampled at centroids of slope units and merged with the number of landslides in those units. The dataset is then split into training and testing subsets, with 80% of the data used for training and 20% reserved for testing. The logistic regression model is trained on the training data to predict the binary outcome of landslide or no-landslide in the slope unit. This model is fitted using the default settings of Python library Scikit-learn's 'LogisticRegression' class, aiming to learn the probability of landslide occurrences as a function of the PGA values. This yields an area under the receiver operating characteristic (ROC) curve of 0.93. The characteristics of the normal cumulative density function (CDF) are a mean value of 0.506 and a standard deviation of 0.127.

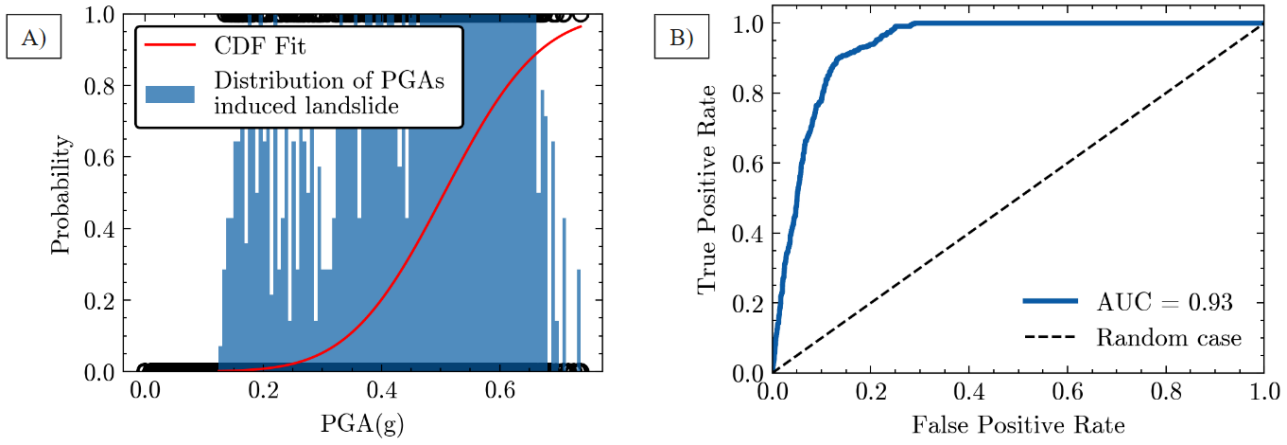


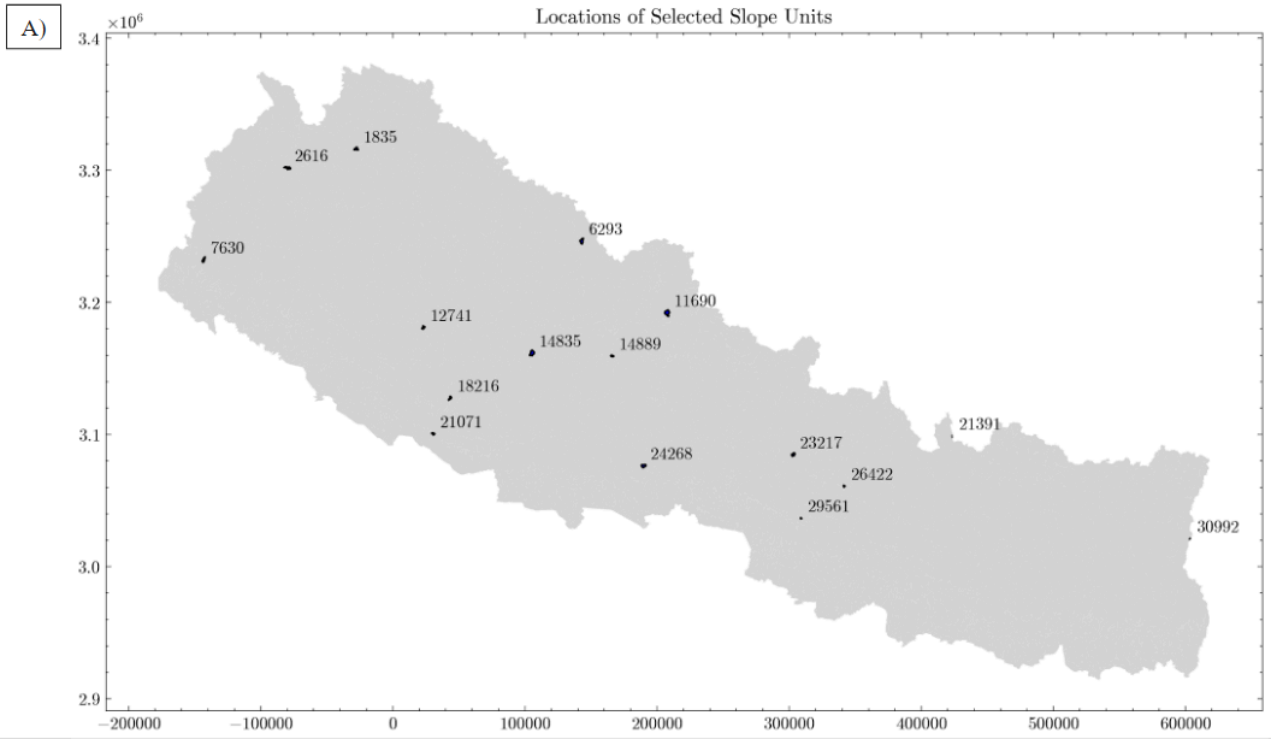
Figure S3: A, relationship between PGA and the probability of coseismic landslide occurrence. B, receiver operating characteristic curve derived for a random selection of 20% of slope units. Value shows the area under the curve (AUC) and the dashed line shows the random case (AUC = 0.5).

Supplemental material 4. Distributions of susceptibility per slope unit

Each slope unit (Fig. S1) contains a unique distribution of landslide susceptibility values. When a slope unit is activated during a model run, sampling this unique distribution to determine whether or not a landslide has been triggered would require the model to hold that distribution in memory, and would slow the model run time substantially. As a simplification, we instead calculate the mean and standard deviation of landslide susceptibility values within each slope unit in advance, and add them as attributes to each slope unit. This allows us to simulate the distribution of susceptibility values with a Gaussian distribution using the same mean and standard deviation during each pass through

the hypergraph network. Values drawn from this synthetic Gaussian distribution are then compared with a uniform random deviate to determine whether one or more landslides has been triggered in that slope unit.

To evaluate the validity of this simplification, Figure S4 shows the actual distributions of landslide susceptibility values from a random selection of slope units across Nepal, along with the corresponding synthetic Gaussian distributions with the same mean and standard deviation values. While the actual distributions take a variety of forms and are both left- and right-skewed, a Gaussian distribution provides a reasonable approximation to the true distribution of values in most cases.



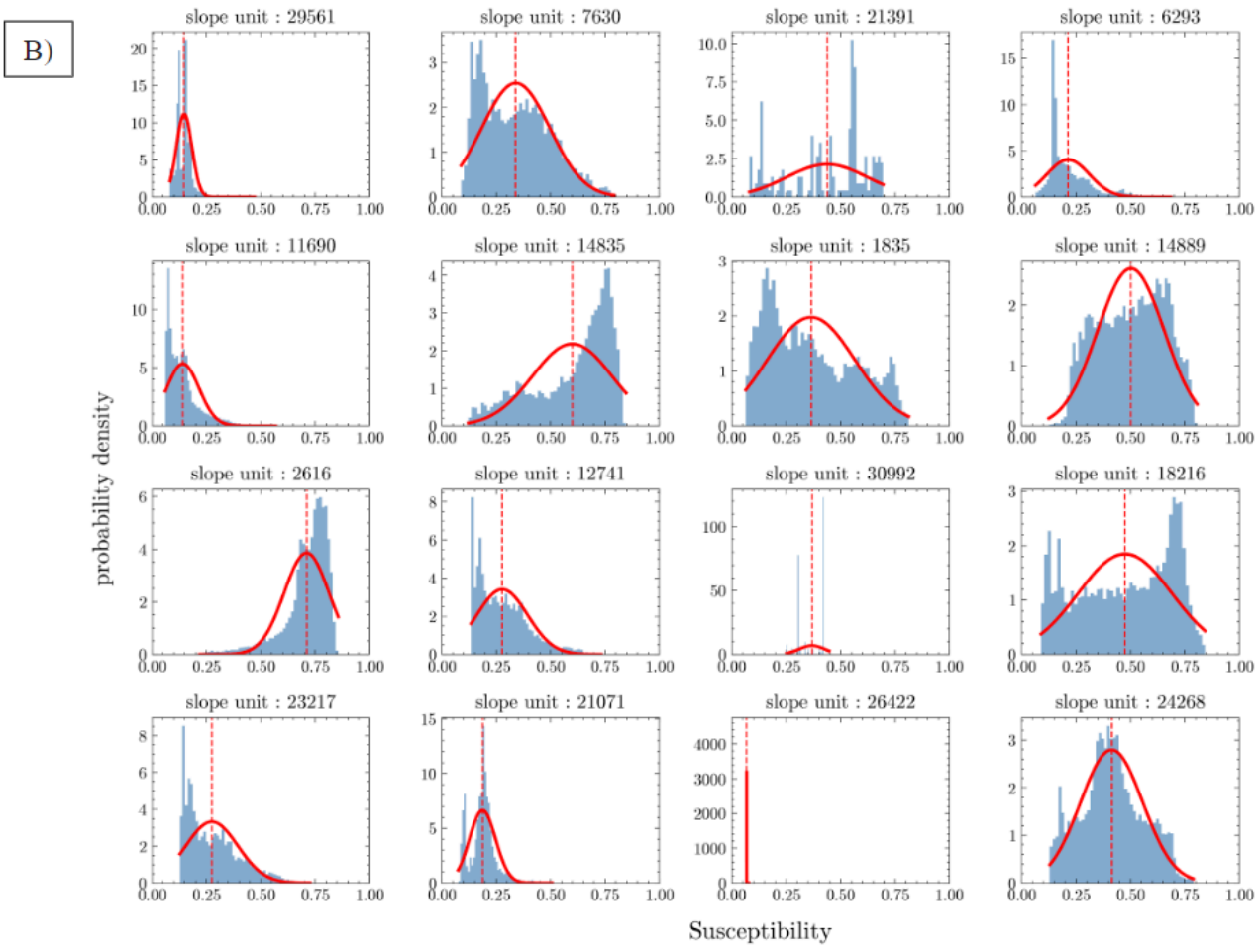


Figure S4: Actual and modelled distributions of landslide susceptibility values in a range of slope units across Nepal. A, locations of randomly-selected slope units. Numbers correspond to sub-figures in panel B. Coordinate axes show easting and northing values in UTM Zone 48. B, actual and modelled synthetic distributions of landslide susceptibility values. Blue bars show actual values as probability density of susceptibility pixel values, red lines show modelled Gaussian distributions with the same mean and standard deviation, and vertical dashed red lines show mean values.

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016) Automatic delineation of geomorphological slope units with r.slopeunits v1. 0 and their optimization for landslide susceptibility modeling. Geoscientific Model Development, 9, 3975-3991, doi:10.5194/gmd-9-3975-2016.

Alvioli, M., Guzzetti, F., & Marchesini, I. (2020) Parameter-free delineation of slope units and terrain subdivision of Italy. *Geomorphology*, 358, 107124, doi:10.1016/j.geomorph.2020.107124.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Kincey, M.E., Rosser, N.J., Robinson, T.R., Densmore, A.L., Shrestha, R., Singh Pujara, D., Oven, K.J., Williams, J.G., & Swirad, Z.M. (2021) Evolution of coseismic and post-seismic landsliding after the 2015 M_w 7.8 Gorkha earthquake, Nepal. *Journal of Geophysical Research-Earth Surface*, doi:10.1029/2020JF005803.

Kincey, M.E., Rosser, N.J., Densmore, A.L., Robinson, T.R., Shrestha, R., Pujara, D.S., Horton, P., Swirad, Z.M., Oven, K.J., & Arrell, K. (2022) Modelling post-earthquake cascading hazards: changing patterns of landslide runout following the 2015 Gorkha earthquake, Nepal. *Earth Surface Processes and Landforms*, 48, 537-554, doi:10.1002/esp.5501.