Explaining the spatial pattern of debris flow and flood hazard in High Mountain Asia

Varvara Bazilova, Tjalling de Haas, Walter Immerzeel

Image source: Google Earth

debris flow dominated

fluvial flow dominated

Original Paper | Published: 19 October 2020

Morphometrical analysis of torrential flows-prone catchments in tropical and mountainous terrain of the Colombian Andes by machine learning techniques

María Isabel Arango, Edier Aristizábal & Federico Gómez

Natural Hazards **105**, 983–1012 (2021) | Cite this article

Open Access | Article

Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China

by Yonghong Zhang [1], Taotao Ge [1], Wei Tian [2,*] and Yuei-An Liou [3,*]

Article | Open Access | Published: 29 August 2019

Assessing Susceptibility of Debris Flow in Southwest China Using Gradient Boosting Machine

Baofeng Di, Hanyue Zhang, Yongyao Liu, Jierui Li, Ningsheng Chen, Constantine A. Stamatopoulos, Yuzhou Luo & Yu Zhan

Scientific Reports **9**, Article number: 12532 (2019) | Cite this article

ELSEVIER

On predicting debris flows in arid mountain belts

Amelie Stolle, Maria Langer, Jan Henrik Blöthe [1], Oliver Korup

Open Access | Article

Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China

by Ke Xiong [1], Basanta Raj Adhikari [1,2], Constantine A. Stamatopoulos [3], Yu Zhan [4], Shaolin Wu [4], Zhongtao Dong [1] and Baofeng Di [1,4,*]

Original Paper | Published: 07 February 2013

Debris-flow susceptibility of upland catchments

Mélanie Bertrand, Frédéric Liébault & Hervé Piégay

Natural Hazards **67**, 497–511 (2013) | Cite this article

What is happening in HMA?

(Glaciers are projected to change (for this amount) (Climate is projected to change (to this amount)

Impact of a global temperature rise of 1.5 degrees Celsius on Asia's glaciers (P. Kraaijenbrink et al., 2017)

- Can we build a machine learning classifier to determine the probability of debris-flow vs flood dominated system?
- What parameters matter for the classification? How does adding climatic features affects the classification?
- What are the regional differences of the debris flow and flood probability?
- Make projections based on climate scenarios

Morphometric + climate

Morphometric

- x_centroid
- y_centroid
- area_m
- perimeter
- mean_elevation
- median_elevation
- std_elevation
- min_elevation
- max_elevation
- range_elevation (relief)
- variance_elevation
- mean_slope
- median_slope
- std_slope
- min_slope
- max_slope
- range_slope
- variance_slope
- Melton_ratio ($relief*area^{0.5}$)
- circularity_ratio
- compactness_coefficient
- **region**

- mean_annual_temp
- mean_jan_temp
- mean_jury_temp
- mean_monsoon_temp
- mean_outside_monsoon_temp
- temp_crosses_zero (frost cracking)
- belowzero_fraction_of_year
- mean_daily_precipitation
- mean_annual_sum_precipitation
- mean_daylymonsoon_precipitation
- mean_monsoon_sum_precipitation
- monsoon_precipitation_fraction
- n_rainy_days (>10mm)
- rainy_days_fraction
- **avgtemp_belowzero**

- glacier_area_sum
- glacier_area_fraction
- **glacier**

- isolated_permafrost_area
- sporadic_permafrost_area
- discontinious_permafrost_area
- continuous_permafrost_area
- sporadic_permafrost_frac
- discontinious_permafrost_frac
- isolated_permafrost_frac
- continious_permafrost_frac
- all_permafrost_frac
- **cont_permafrost_frac > 50%**
- **any_permafrost**

Gael Varoquaux
@GaelVaroquaux

For thousands of data points and moderate dimensionality (99% of cases), gradient-boosted trees provide the necessary regression model scikit-learn.org/stable/modules...
They are robust to data distribution and support missing values (even outside MAR* settings arxiv.org/abs/1902.06931)

* MAR = Missing At Random

Gradient boosting algorithm (explained) + catboost

GB algorithm

Gradient boosting algorithm (explained) + catboost

GB algorithm

(Same slide but add catboost)

Building the model: how good is it?

Morphometric

- Accuracy : 91 %

(fraction of correct predictions)

- Confusion matrix: [145., 28.]
  [17., 522.]

Morphometric + climate

- Accuracy: 92 %
- Confusion matrix: [148., 25.]
  [14., 525.]

TP	FN
FP	TN

- auc = area under curve
- only “tuned” parameter: number of trees (iterations)
- debris flow (1): 539, flood (0): 173
- accuracy, when guessing randomly: 75.7 %

(Here example maps with “predicted” catchments): same area but 2 different “models” we built

Why does Catboost model make this predictions?

Morphometric

Morphometric + climate

(Here should be smth about partial dependancy, but I have to think a bit more about it)

Conclusions

- We can build a machine learning classifier for distinguishing debris-flow dominated systems from flood dominated ones
- Climate data adds a lot of information to the model, but (all other things being equal) does not improve model performance
-

Outlook

- Extend the dataset for “creating” the model by covering more diverse regions
- Add vegetation cover to the feature list
- Apply the model to the “new” areas (i.e. catchments without alluvial fan)
- To see the effect of the climate change - use RCP scenarios as a climate information