



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

Varya Bazilova,

Tjalling de Haas, Walter Immerzeel



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



Varya Bazilova,
Tjalling de Haas, Walter Immerzeel



Dorje Dolma lama
@DolmaLama444

...

Over 50 people are missing in the Melamchi and Indrawati rivers' flooding. The floods have also caused damages to the dam in Melamchi drinking water project, Timbu Bazaar, Chanaute Bazaar, Talamarang Bazaar and Melamchi Bazar.



8:37 PM · Jun 16, 2021 · Twitter for Android



Dorje Dolma lama
@DolmaLama444

...

Over 50 people are missing in the Melamchi and Indrawati rivers' flooding. The floods have also caused damages to the dam in Melamchi drinking water project, Timbu Bazaar, Chanaute Bazaar, Talamarang Bazaar and Melamchi Bazar.



Kaushal Gnyawali
@KaushalGnyawali

...

Landslide dam outburst seems to have Melamchi flooding in Sindhupalchowk, Nepal





Dorje Dolma lama
@DolmaLama444

Over 50 people are missing in the Melamchi and Indrawati rivers' flooding. The floods have also caused damages to the dam in Melamchi drinking water project, Timbu Bazaar, Chanaute Bazaar, Talamarang Bazaar and Melamchi Bazar.



Kaushal Gnyawali
@KaushalGnyawali

Landslide dam outburst seems to have Melamchi flooding in Sindhupalchowk, Nepal



Colin McCarthy @US_Stormwatch · Aug 30

Hard to comprehend the scale of the **flood** disaster in Pakistan, the 5th most populated nation in the world.

Nearly 1400 dead, 1 million houses damaged or destroyed, and 50,000,000 people displaced.

1/3 of the country is underwater.



▶ 2.8M views

0:03 / 0:24 ⏴ ↗



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](#)

Article | Open Access | Published: 29 August 2019

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

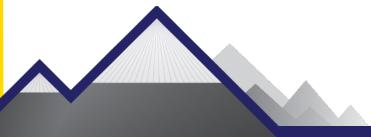
[Baofeng Di](#), [Hanyue Zhang](#), [Yongyao Liu](#), [Jierui Li](#), [Ningsheng Chen](#), [Constantine A. Stamatopoulos](#),
[Yuzhou Luo](#) & [Yu Zhan](#)

[Scientific Reports](#) 9, Article number: 12532 (2019) | [Cite this article](#)

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,*}



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](#)

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,*}

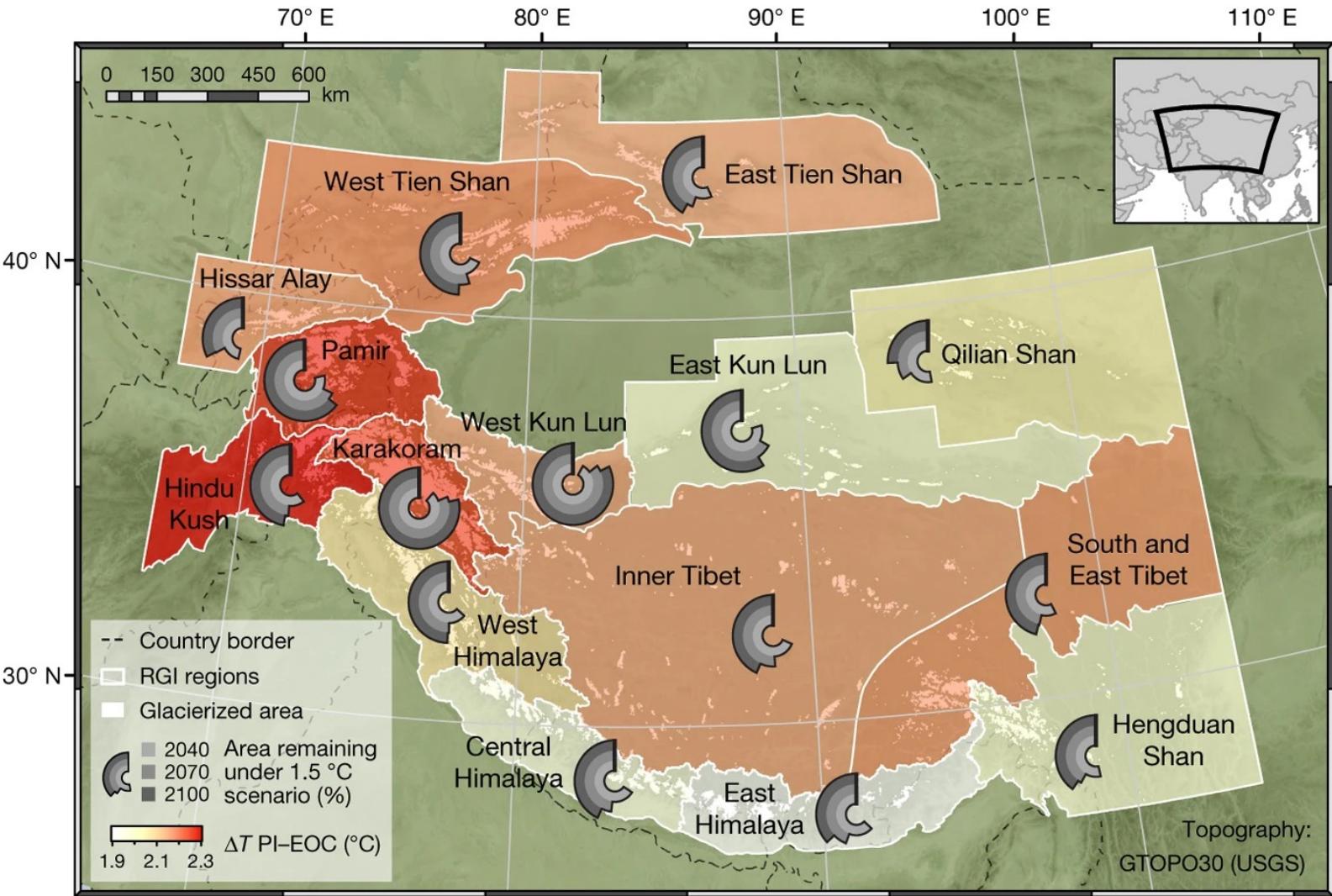
ELSEVIER

On predicting debris flows in arid mountain belts

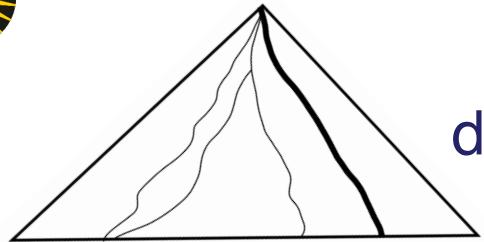
[Amelie Stolle](#) [Maria Langer](#) [Jan Henrik Blöthe](#) ¹ [Oliver Korup](#)



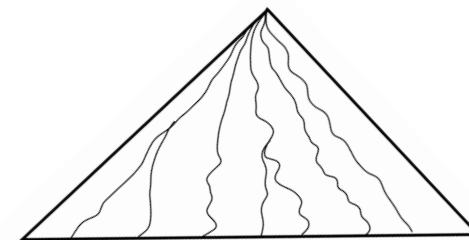
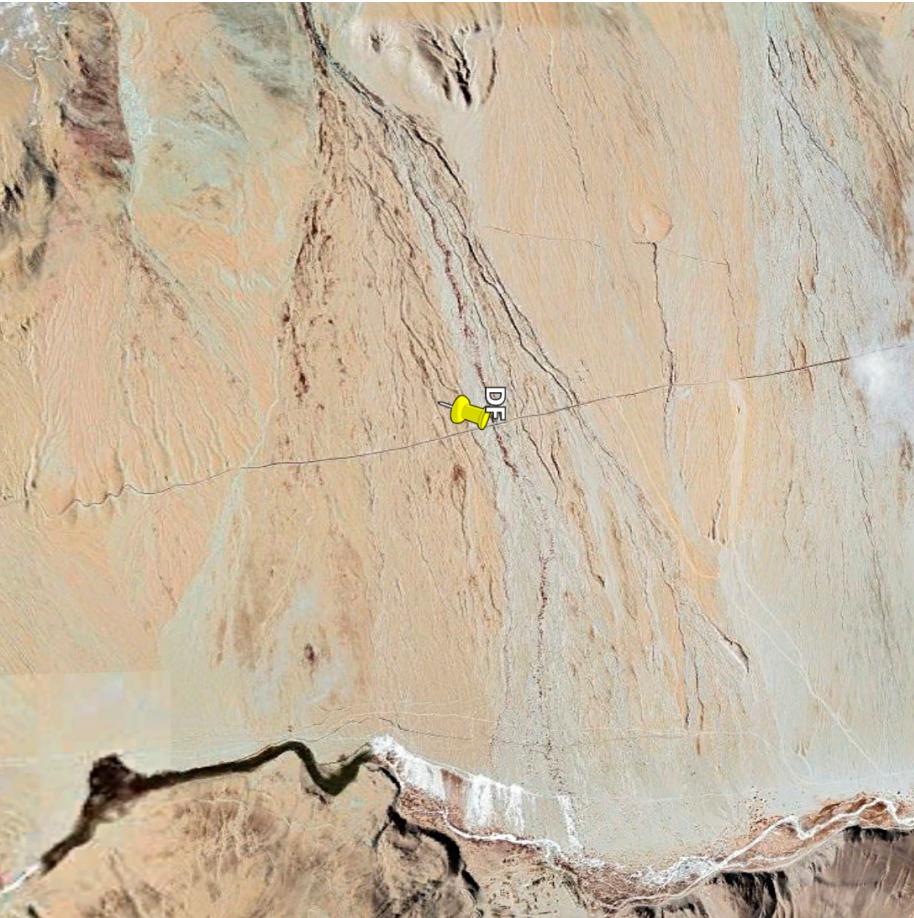
What is happening in HMA?



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



debris flow dominated



fluvial flow dominated

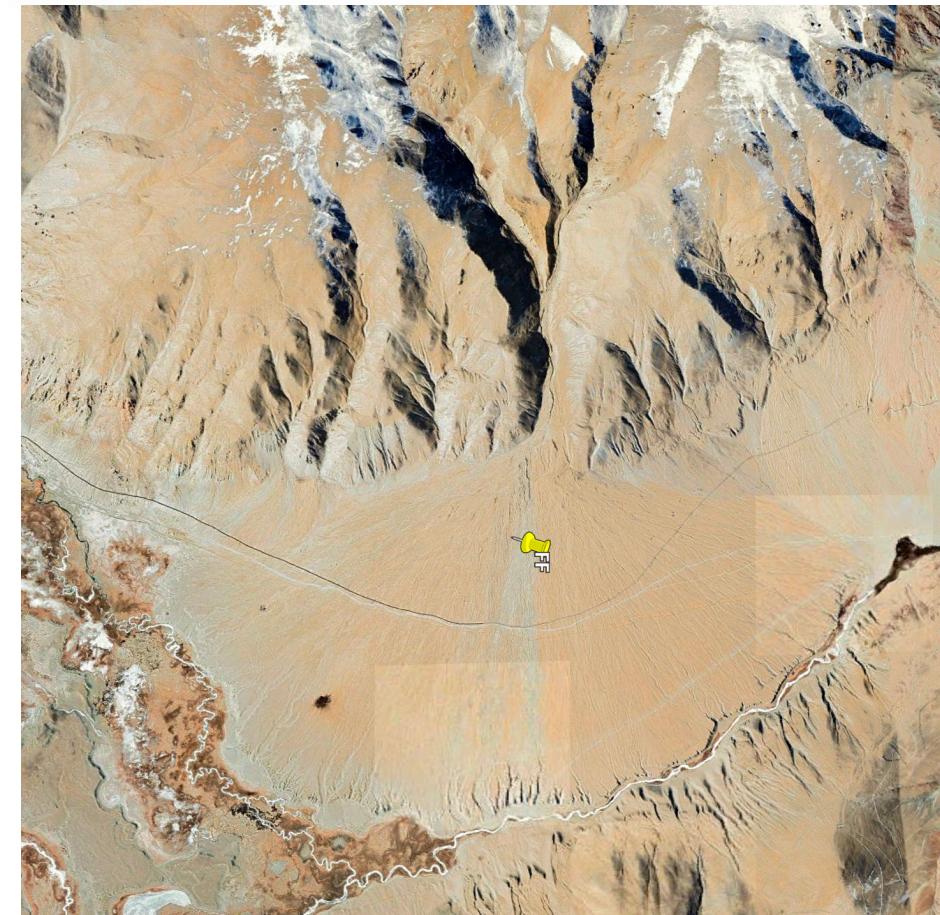


Image source: Google Earth



Research questions

- to use ML classifier to estimate probabilities of debris-flow vs flood dominated system
- to identify the parameters, that matter for the classification
- to see if adding climatic features affects the classification
- to find out if there are any regional differences
- to make projections based on climate scenarios



frame the
question

get data
clean up

split data:
train & test

machine learning:
build the model

tune the
model

evaluate
the model

make
predictions



frame the
question

get data
clean up

split data:
train & test

machine learning:
build the model

tune the
model

evaluate
the model

make
predictions



frame the
question

get data
clean up

split data:
train & test



machine learning:
build the model



tune the
model



evaluate
the model



make
predictions



frame the
question

get data
clean up

split data:
train & test

machine learning:
build the model

tune the
model

evaluate
the model

make
predictions





frame the
question

get data
clean up

split data:
train & test

machine learning:
build the model

tune the
model

evaluate
the model

make
predictions





frame the
question



get data
clean up



split data:
train & test



machine learning:
build the model



tune the
model



evaluate
the model



make
predictions



frame the
question



get data
clean up



split data:
train & test



machine learning:
build the model



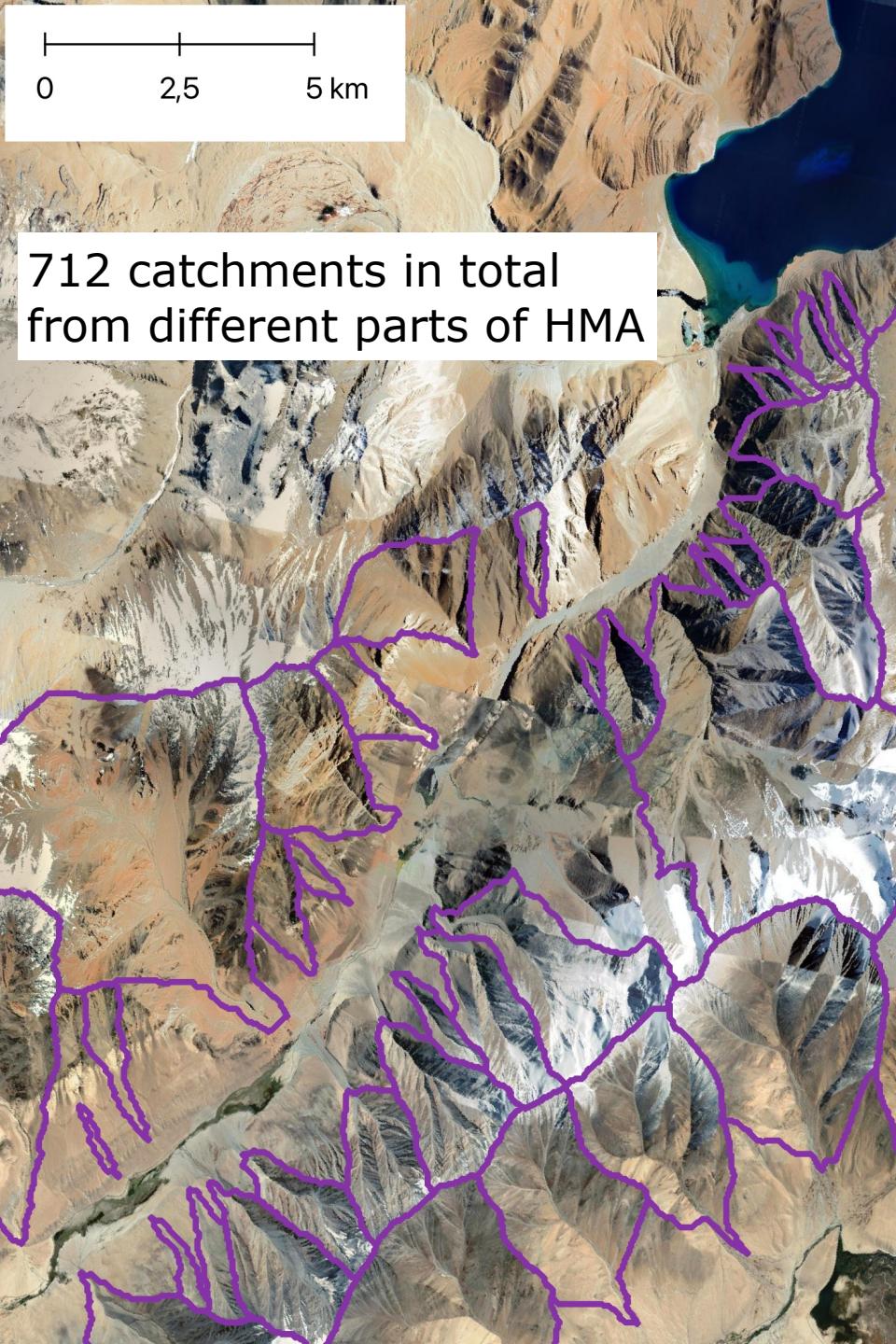
tune the
model



evaluate
the model



make
predictions



Morphometric + climate

- 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

Morphometric

- mean_annual_temp
- mean_jan_temp
- mean_july_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
- discontinuous_permafrost_area
- continuous_permafrost_area
- sporadic_permafrost_frac
- discontinuous_permafrost_frac
- isolated_permafrost_frac
- continuous_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...

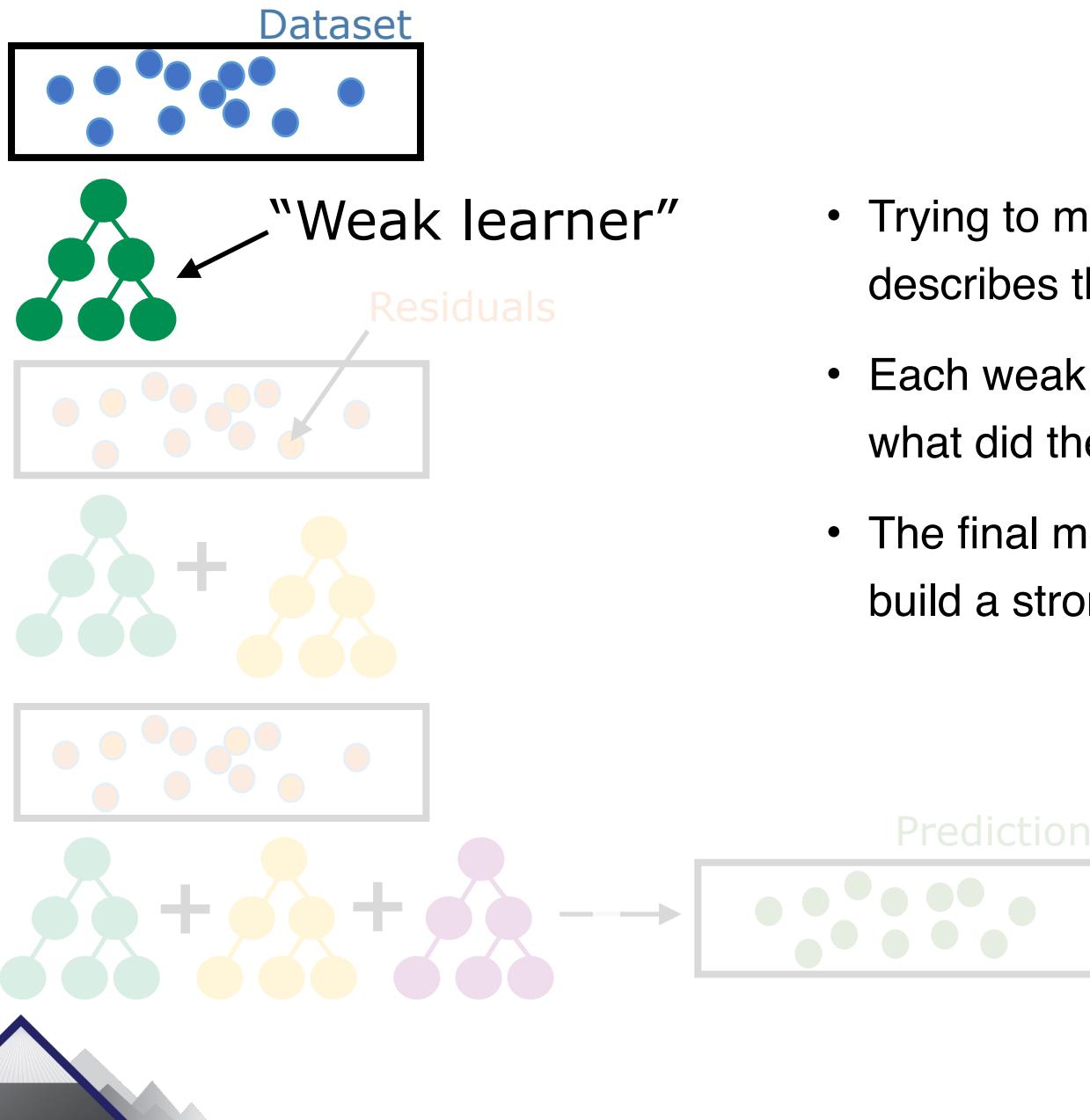
They are robust to data distribution and support missing values (even outside MAR^{*} settings

arxiv.org/abs/1902.06931)

* MAR = Missing At Random



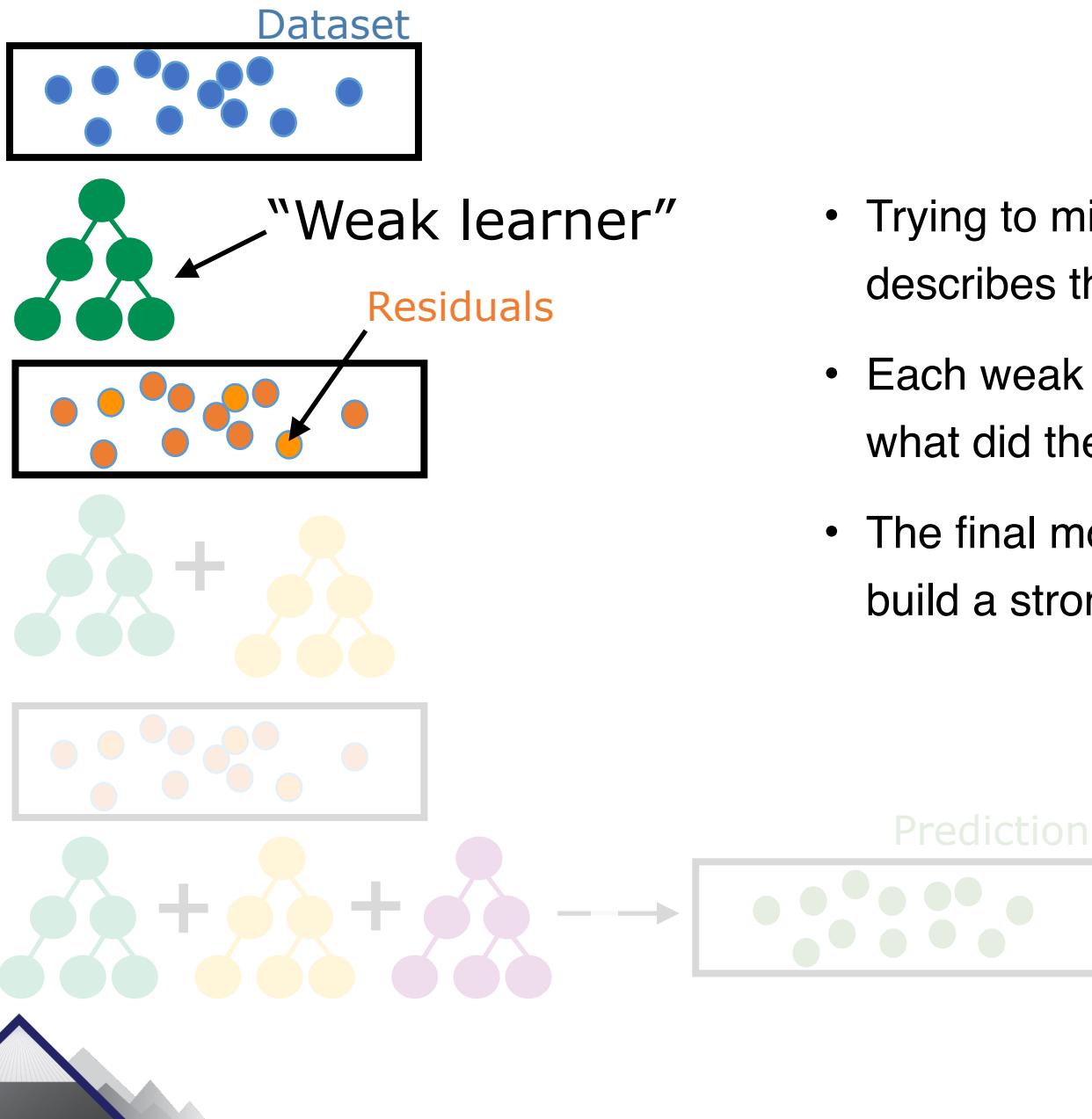
Gradient boosted decision trees



- Trying to minimize the “loss function” (function, that describes the error) on every iteration
- Each weak learner (i.e. tree/iteration) is trying to learn what did the previous one did “wrong” and do better
- The final model is the “combination” of all weak trees to build a strong classifier



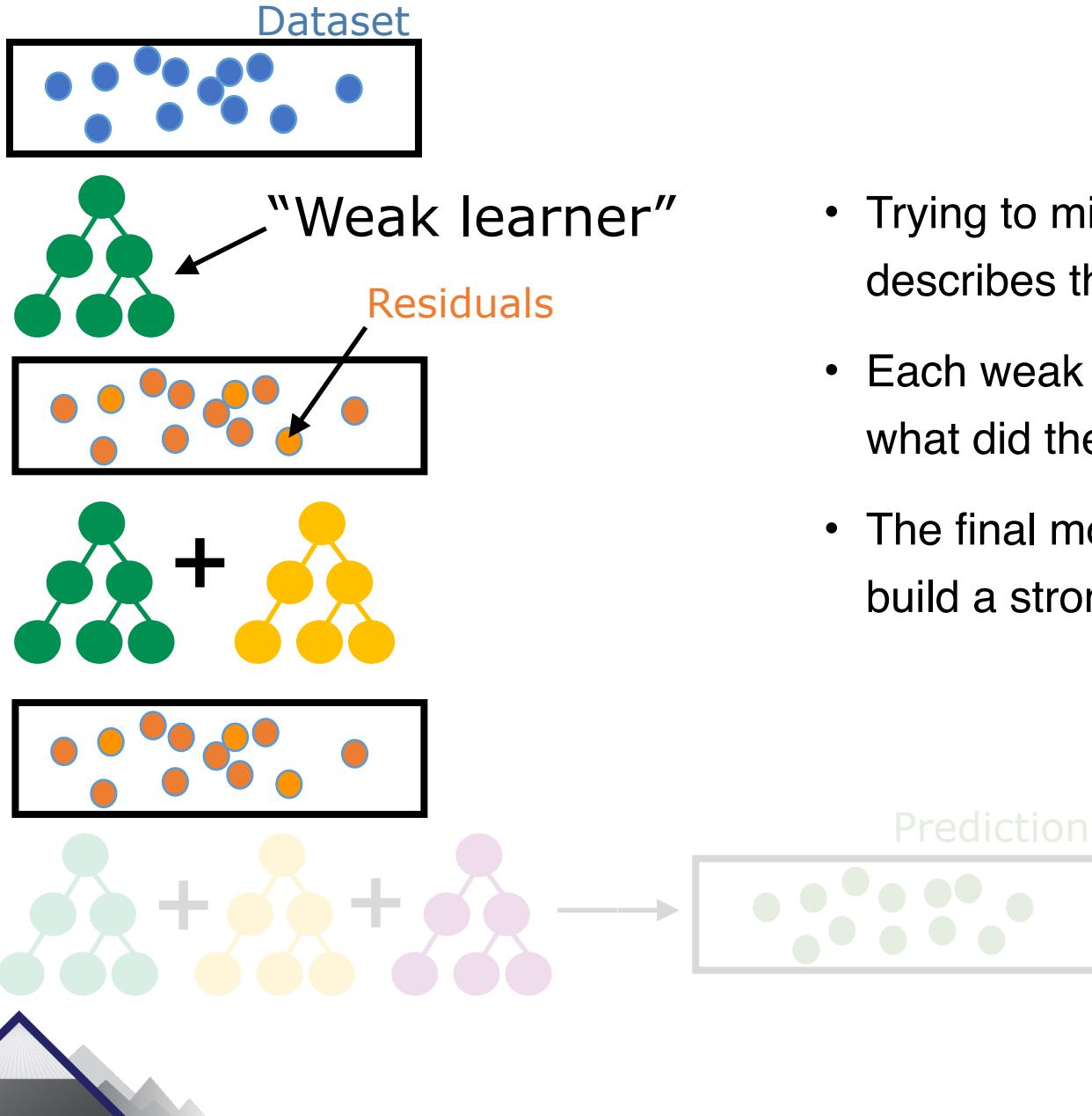
Gradient boosted decision trees



- Trying to minimize the “loss function” (function, that describes the error) on every iteration
- Each weak learner (i.e. tree/iteration) is trying to learn what did the previous one did “wrong” and do better
- The final model is the “combination” of all weak trees to build a strong classifier



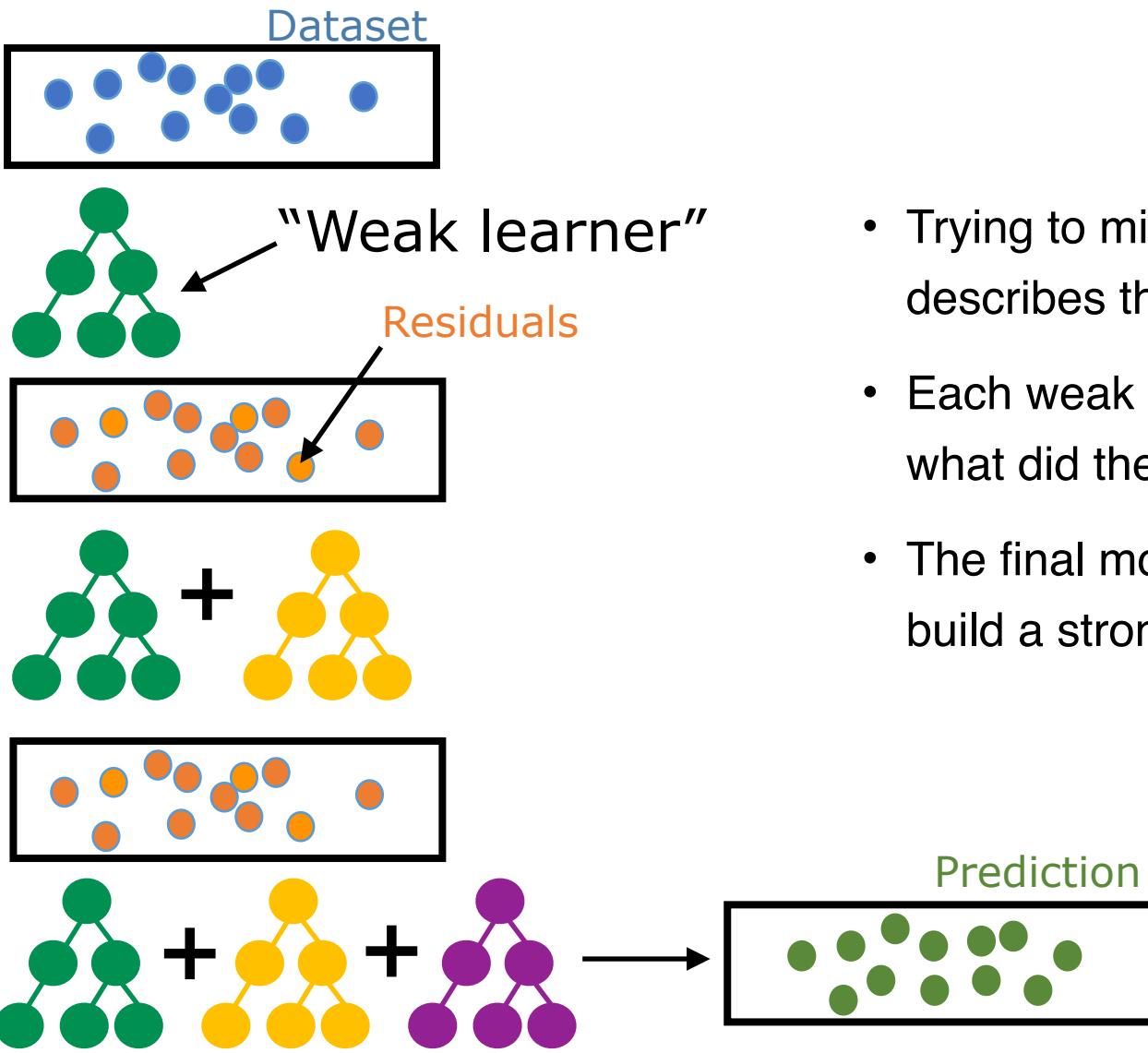
Gradient boosted decision trees



- Trying to minimize the “loss function” (function, that describes the error) on every iteration
- Each weak learner (i.e. tree/iteration) is trying to learn what did the previous one did “wrong” and do better
- The final model is the “combination” of all weak trees to build a strong classifier



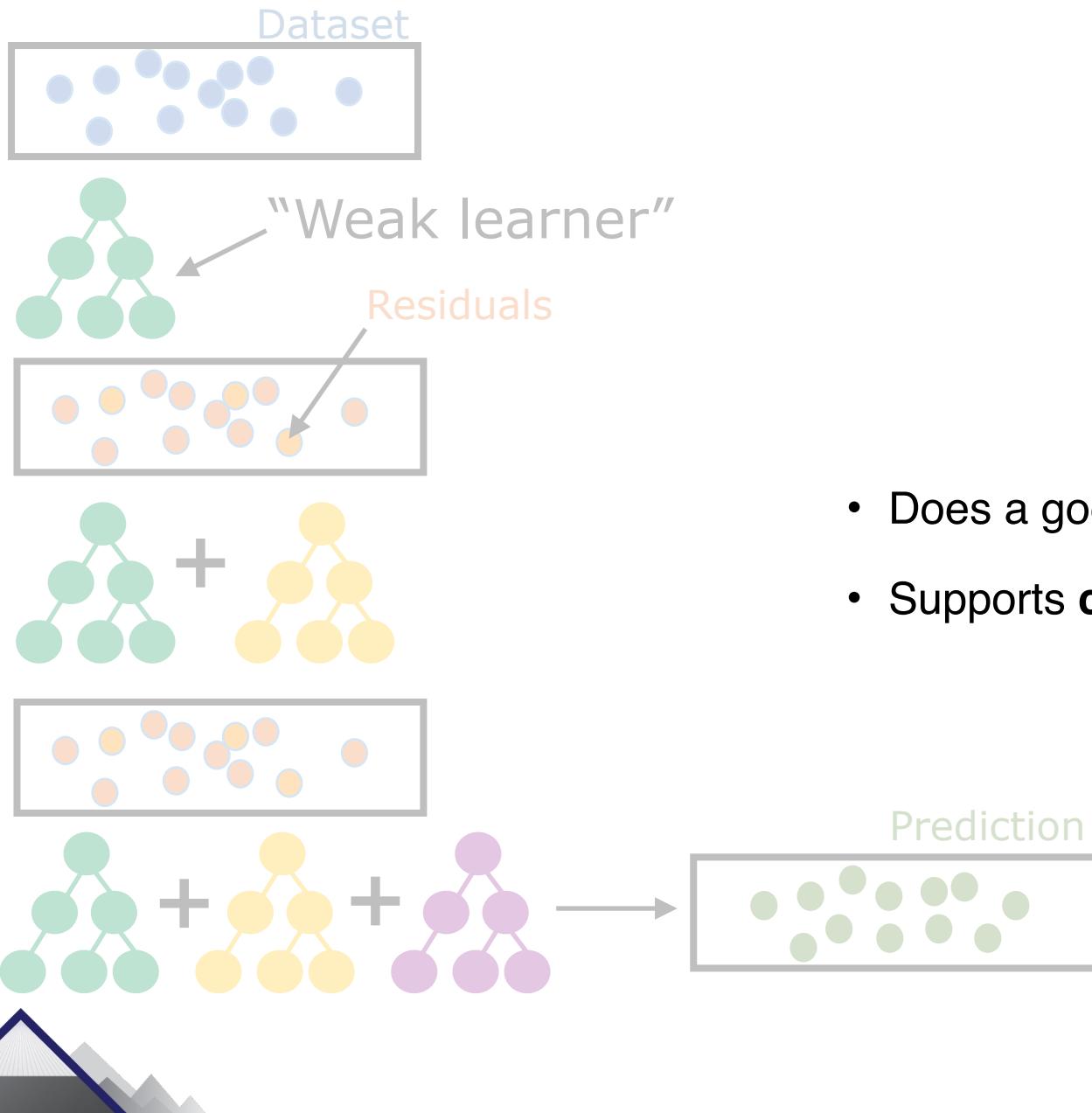
Gradient boosted decision trees



- Trying to minimize the “loss function” (function, that describes the error) on every iteration
- Each weak learner (i.e. tree/iteration) is trying to learn what did the previous one did “wrong” and do better
- The final model is the “combination” of all weak trees to build a strong classifier



Implementation: Catboost



- Does a good job as an “out of the box” tool
- Supports **categorical** features (predictors) as an input



Building the model: how good is it?

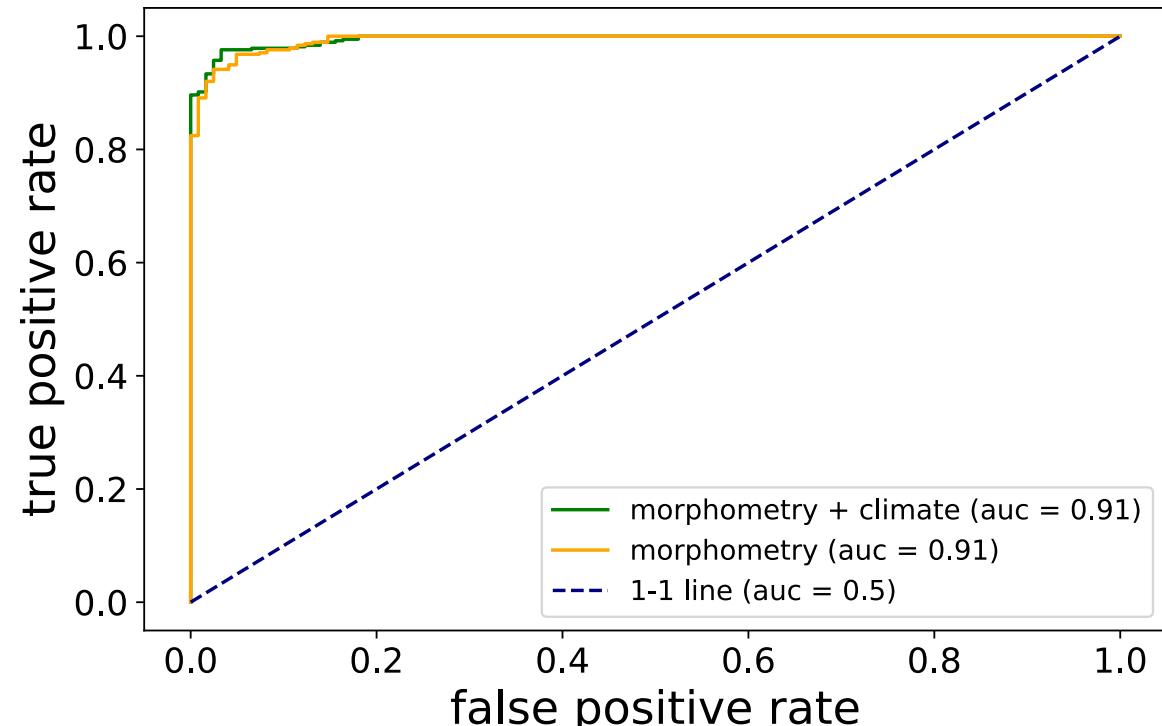
Morphometric

- Accuracy : 91 %
(fraction of correct predictions)
- Confusion matrix: [145., 28.]
[17., 522.]

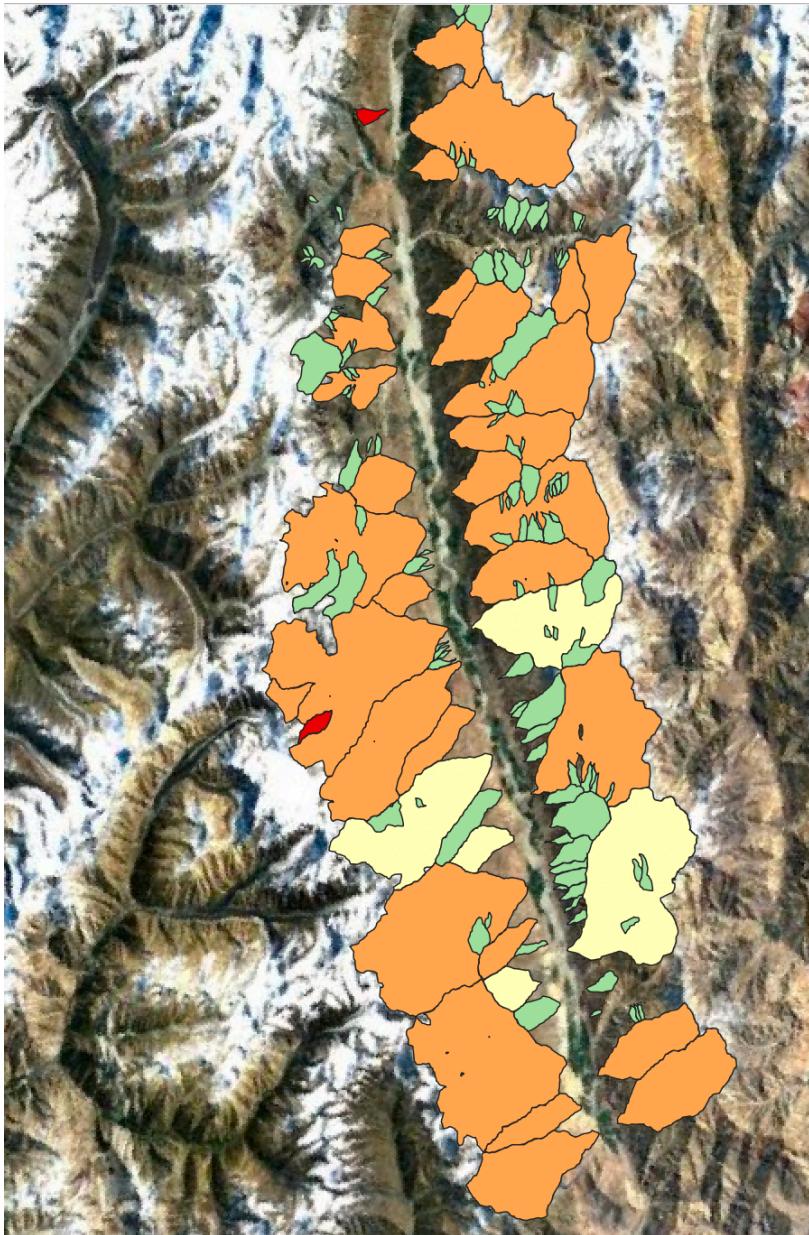
Morphometric + climate

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

TP	FN
FP	TN

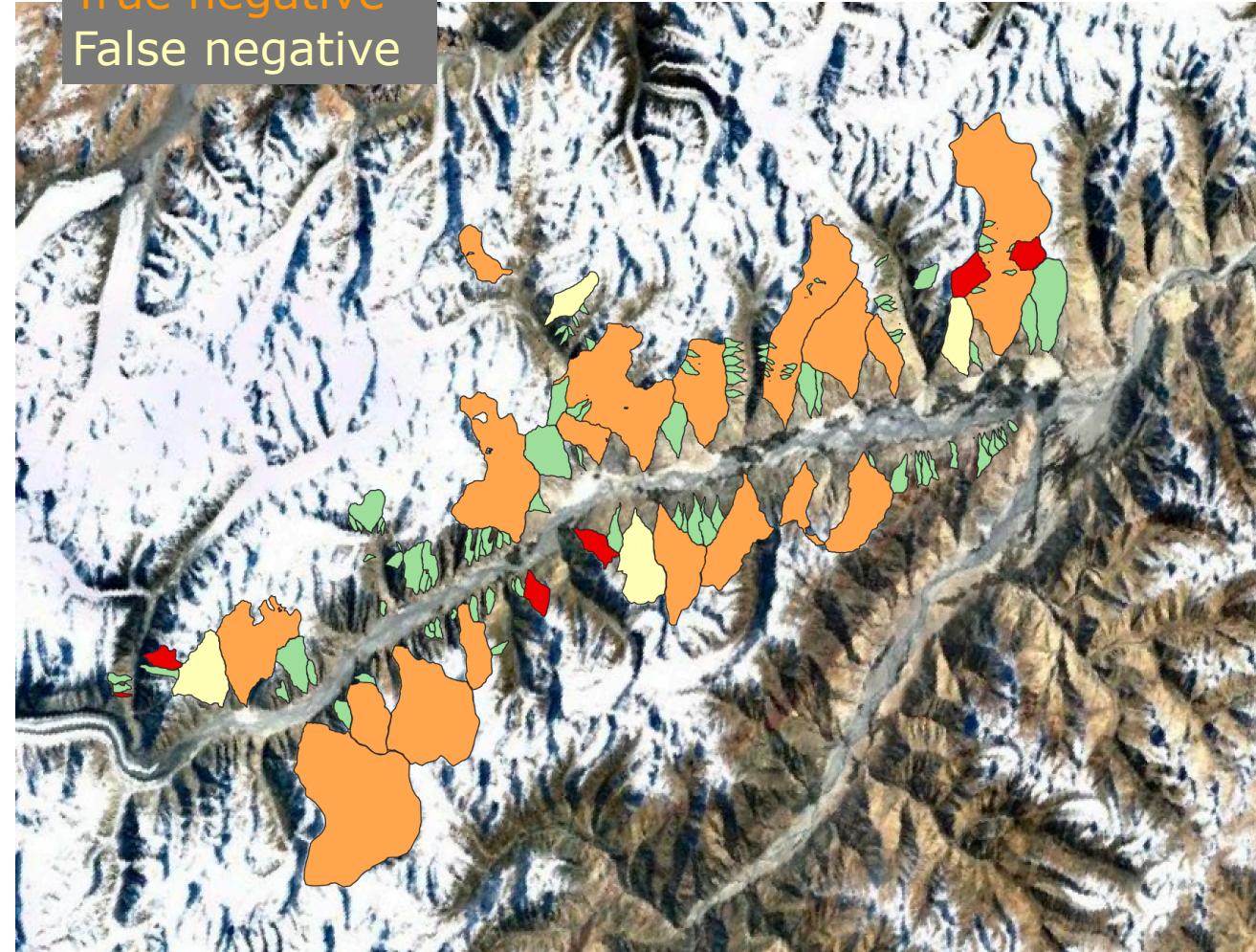


- auc = area under curve
- only “tuned” parameter: number of trees (iterations)
- debris flow (1): 539, flood (0): 173
- accuracy, when guessing randomly: 75.7 %



somewhere in Tajikistan

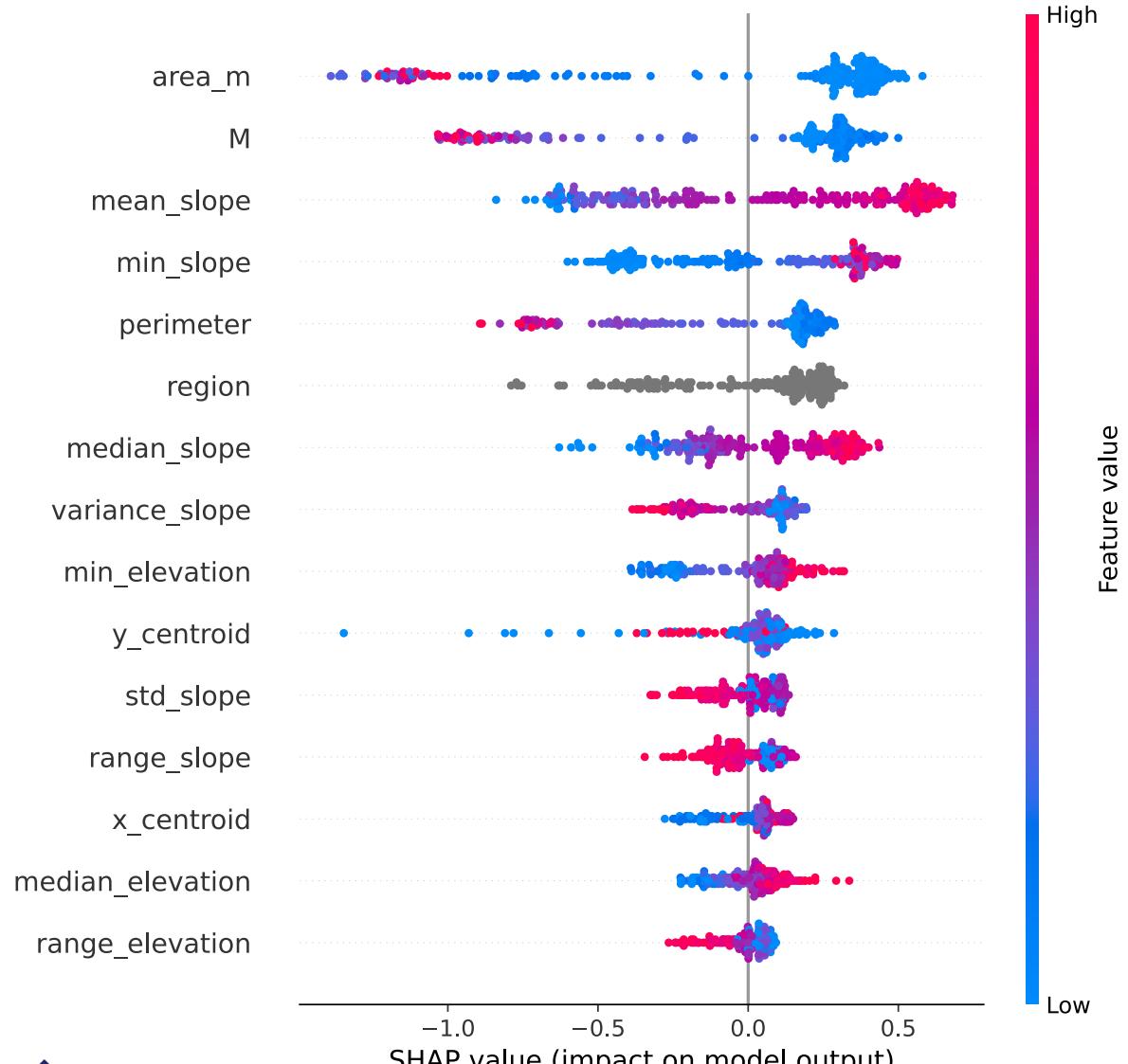
True positive
False positive
True negative
False negative



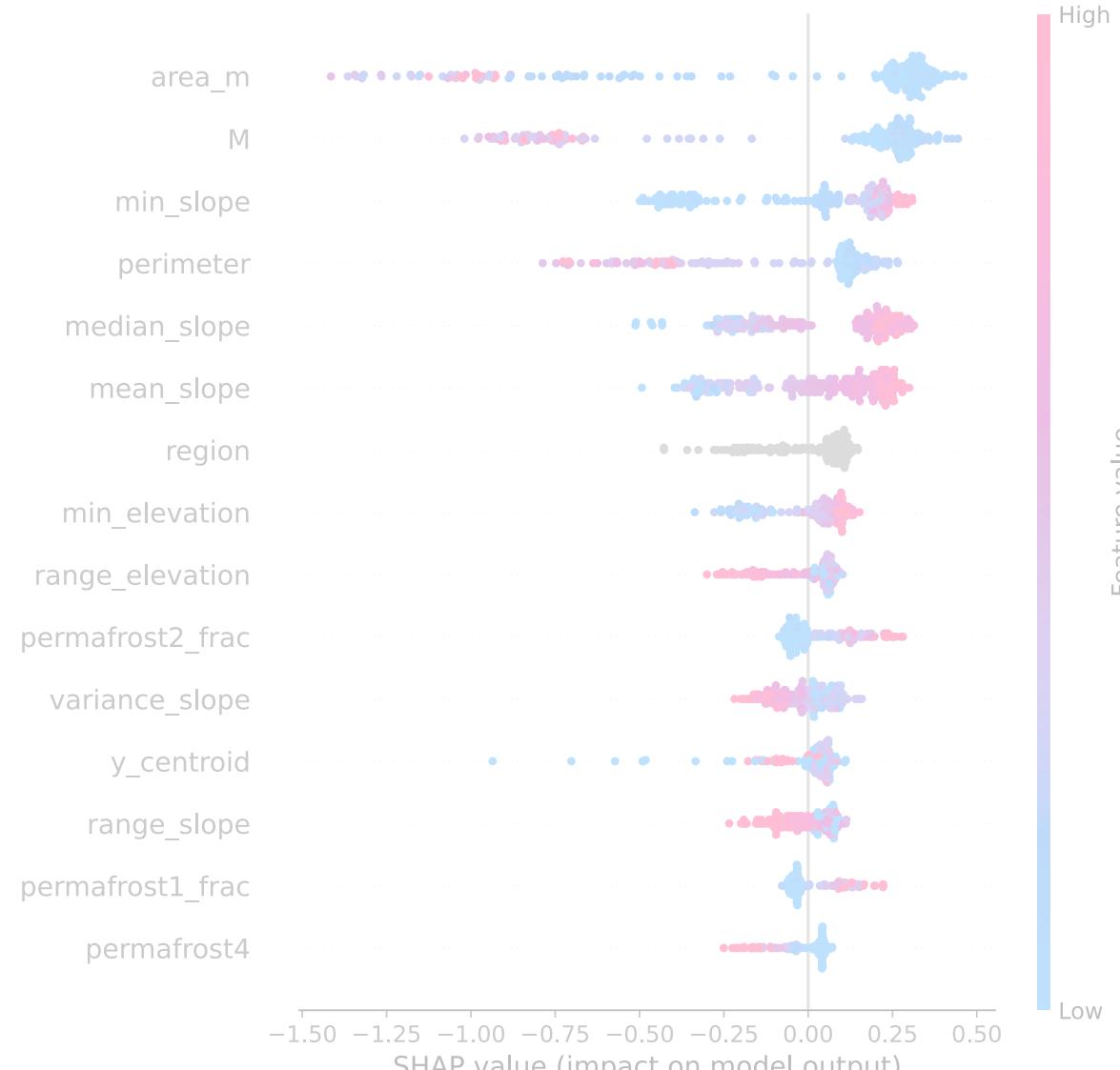
somewhere in Karakoram



Why does Catboost model make this predictions?



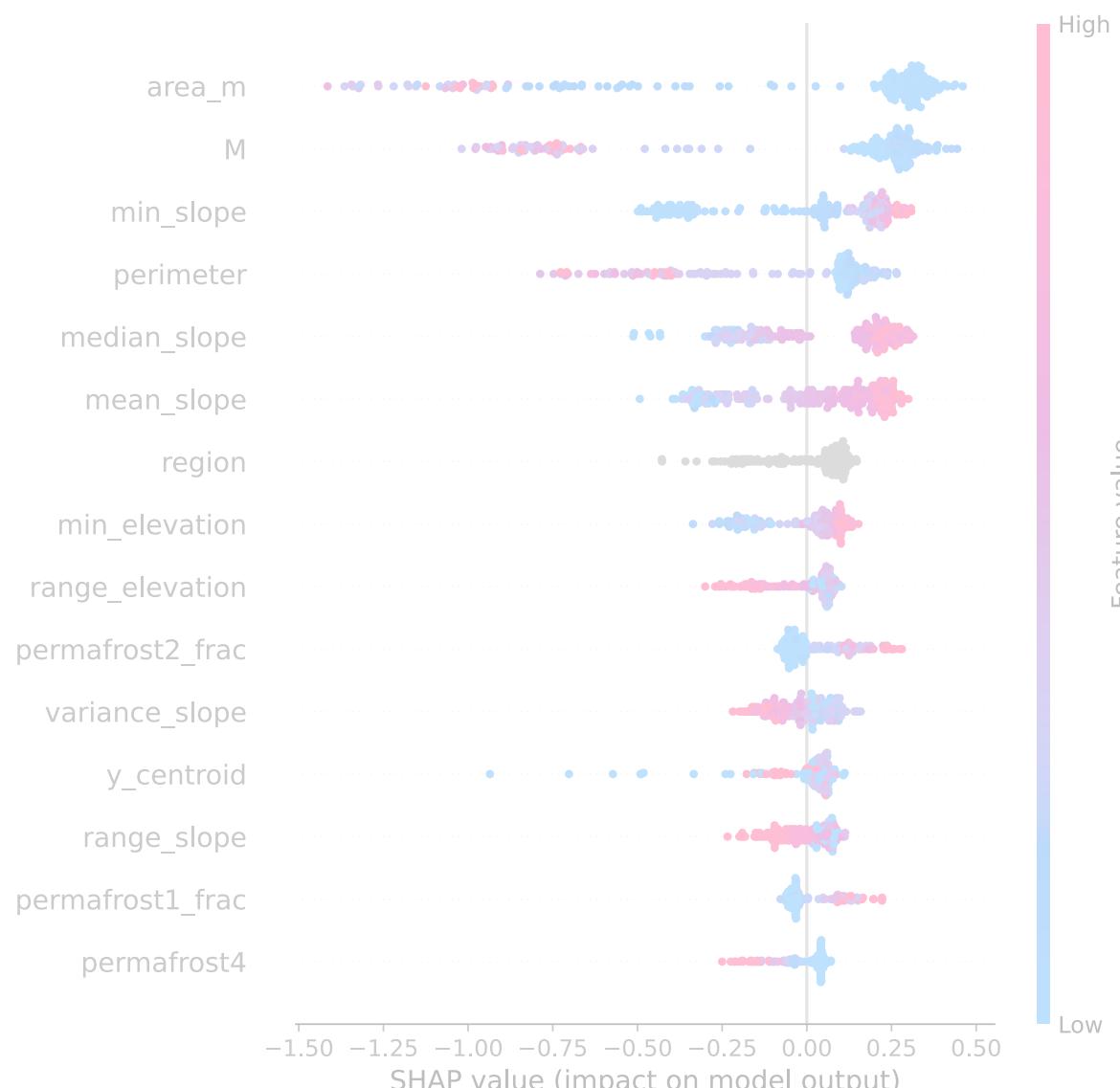
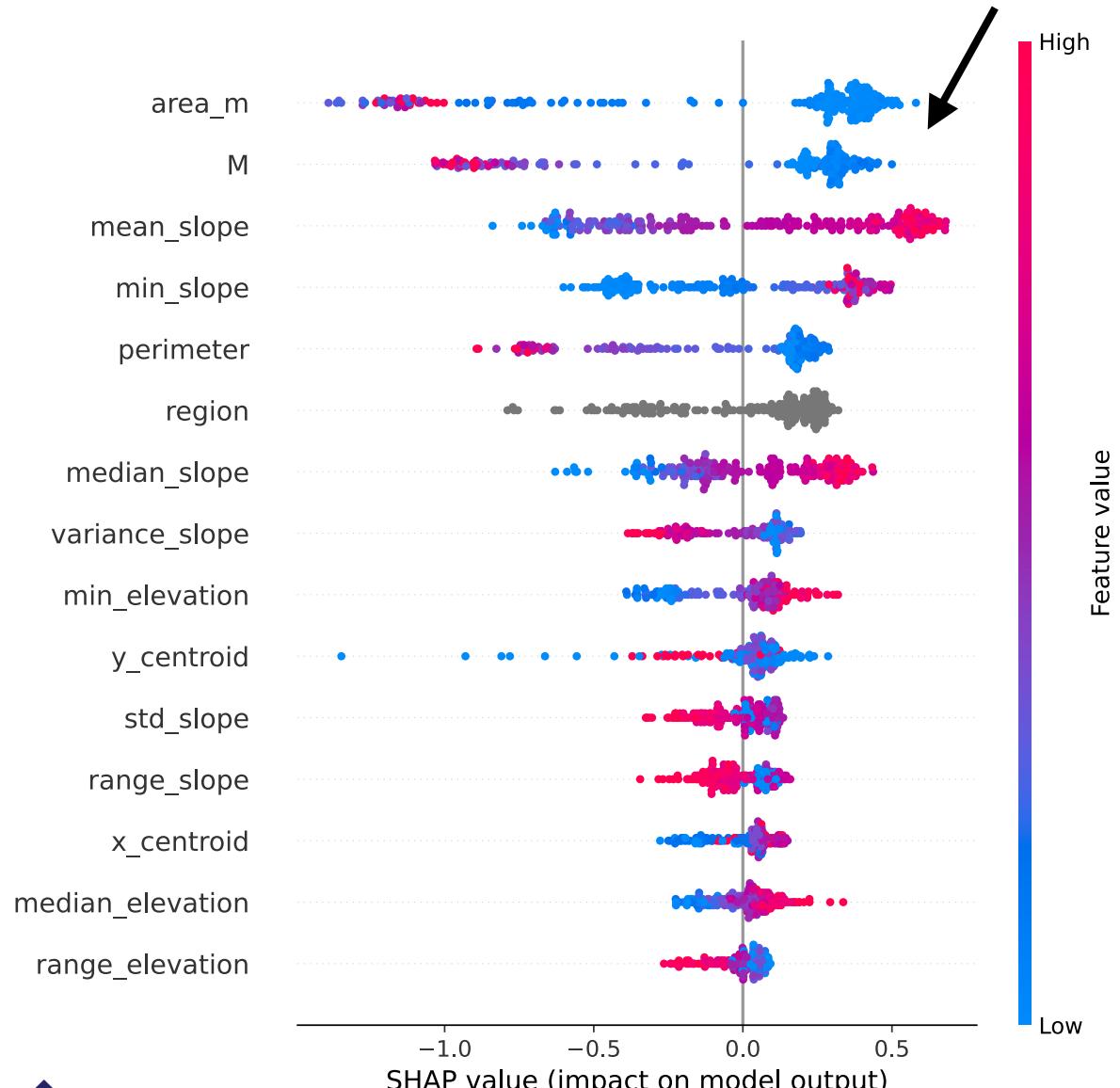
Morphometric



Morphometric + climate

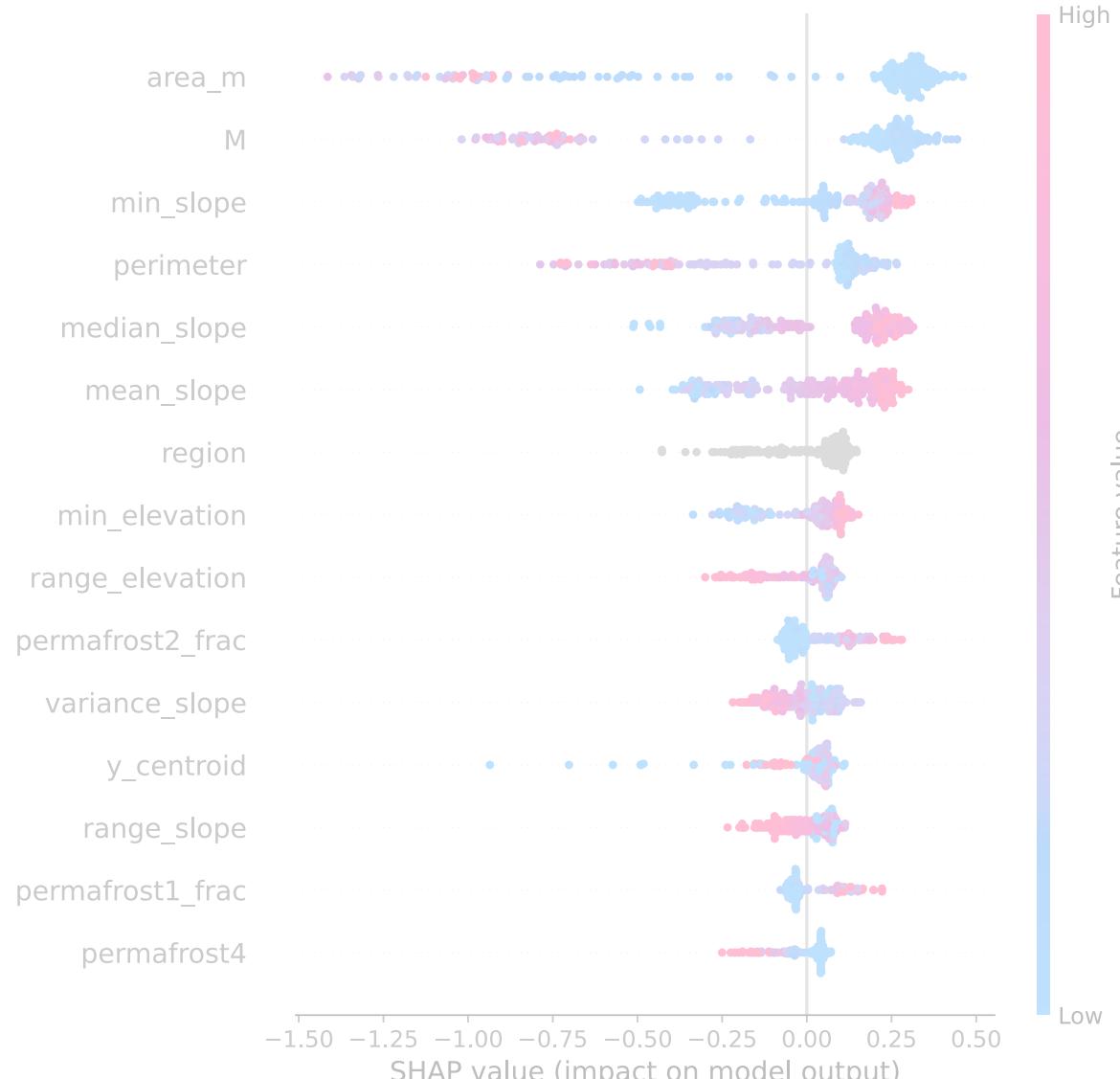
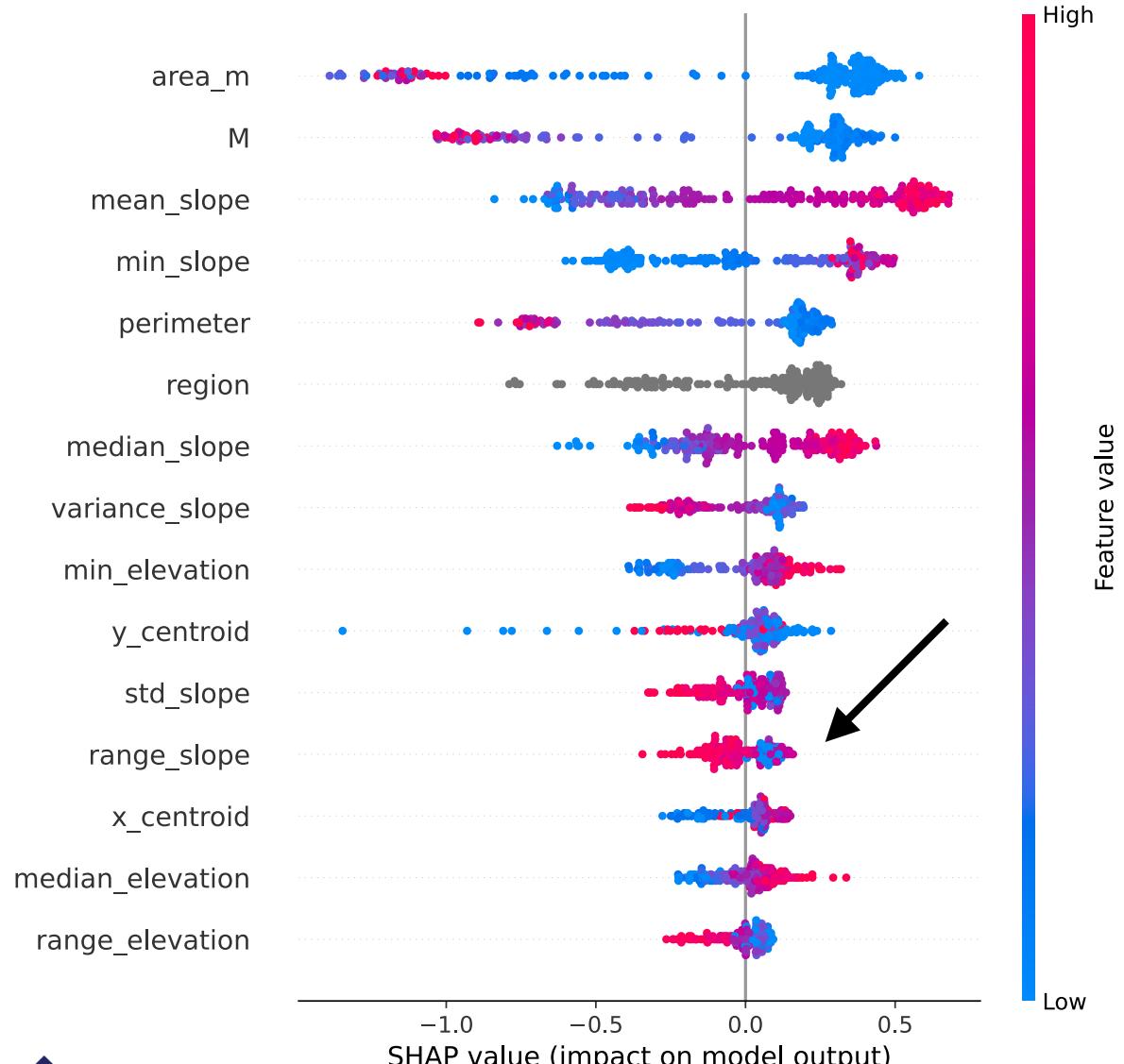


Why does Catboost model make this predictions?



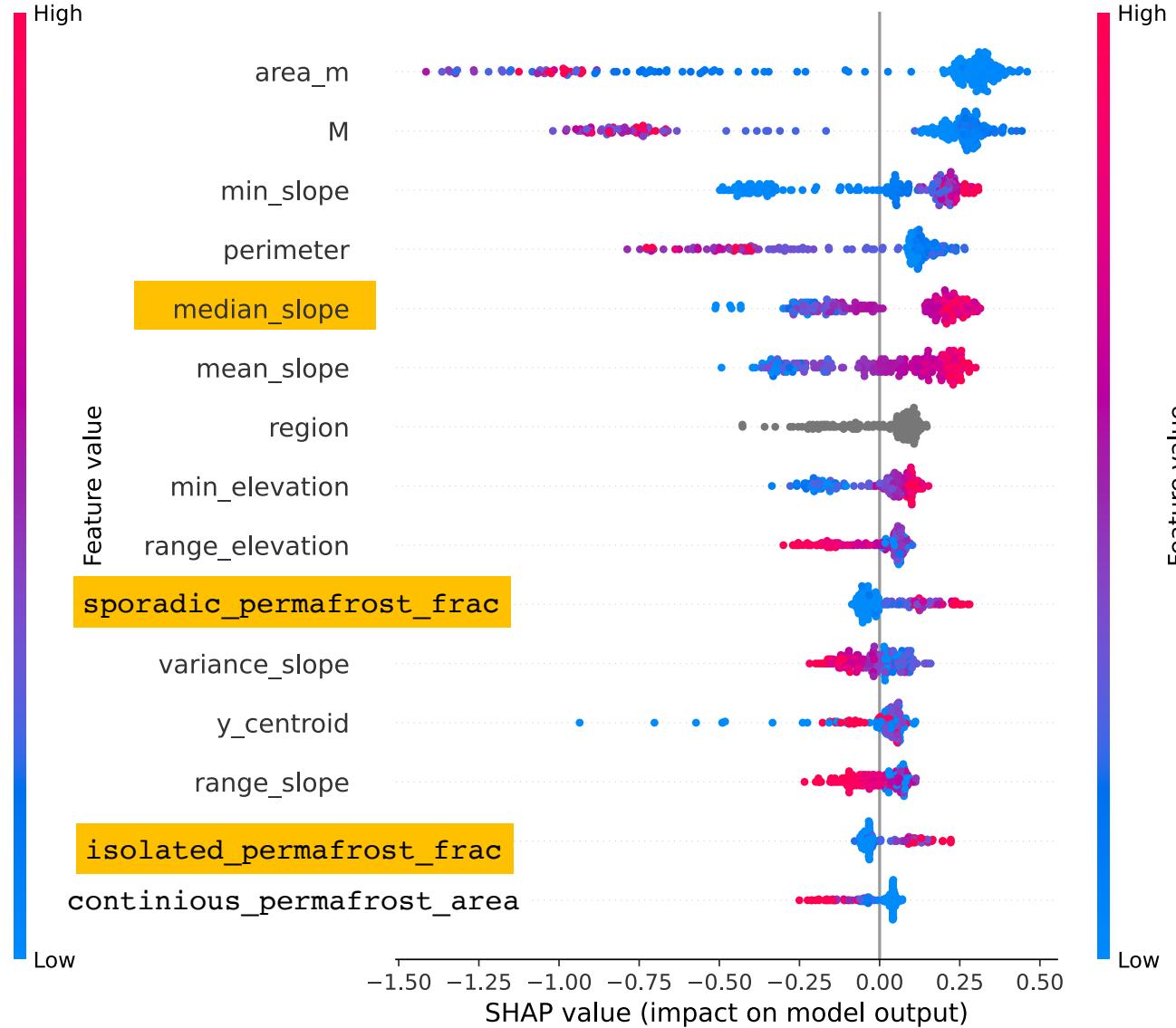
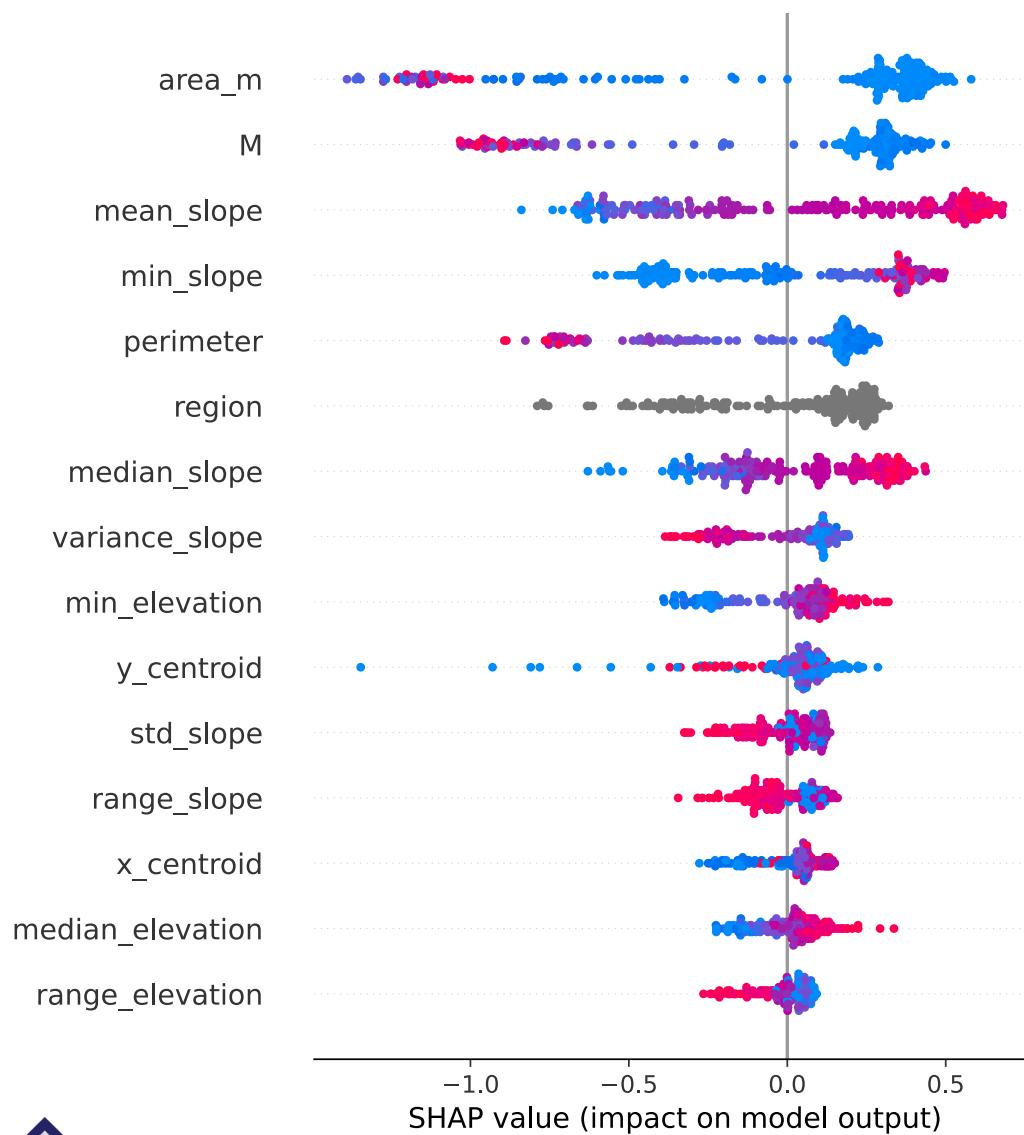


Why does Catboost model make this predictions?





Why does Catboost model make this predictions?



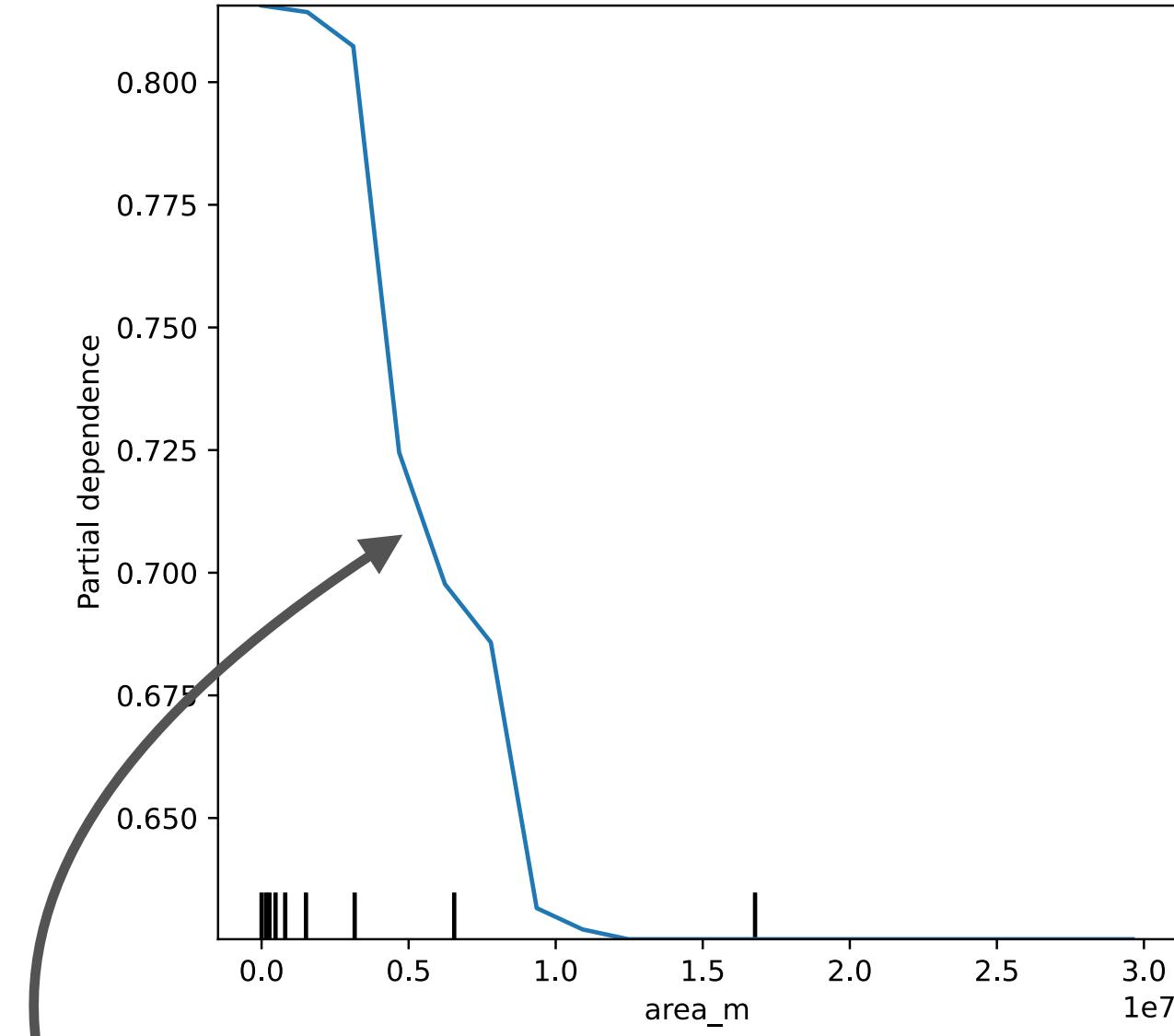


What should we do to the catchment for it to “switch”:

all things being equal, how can we make a “flood” catchment produce “debris flow”



*average probability of a DF (1)
for scenario, where all catchment
have the same area (x-axis)*





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

