



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

**Varvara 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

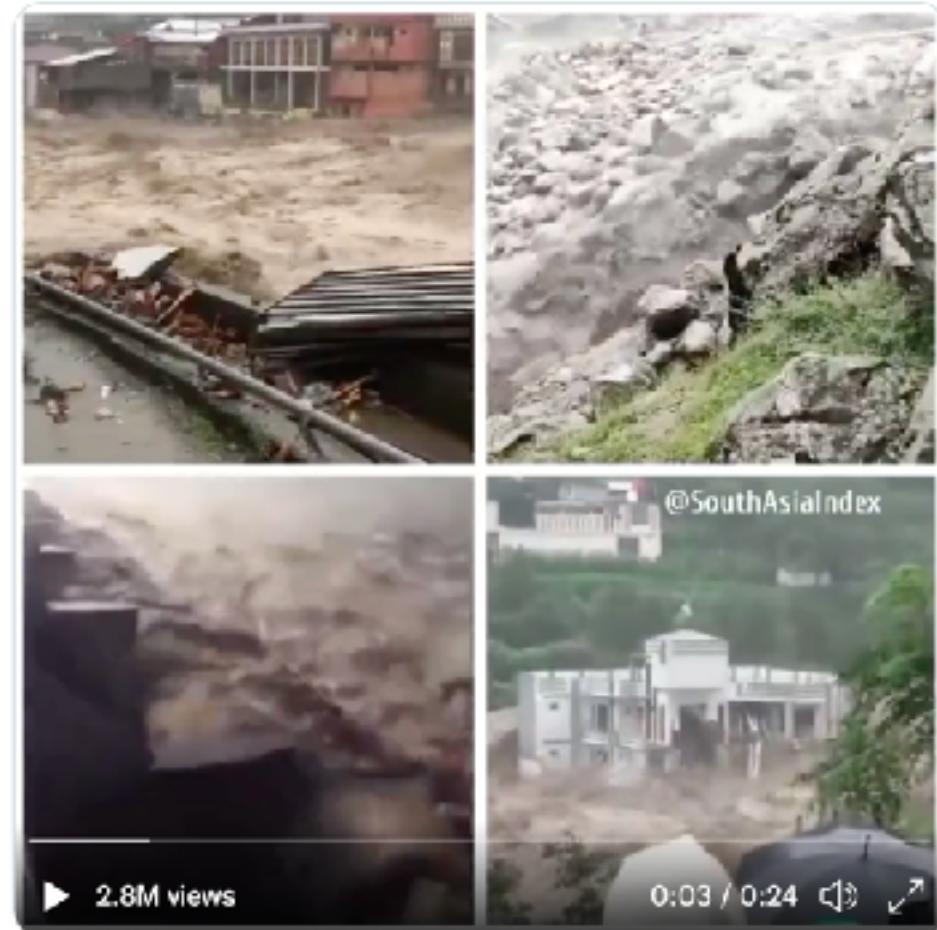


Colin McCarthy @US\_Stormwatch · Aug 30

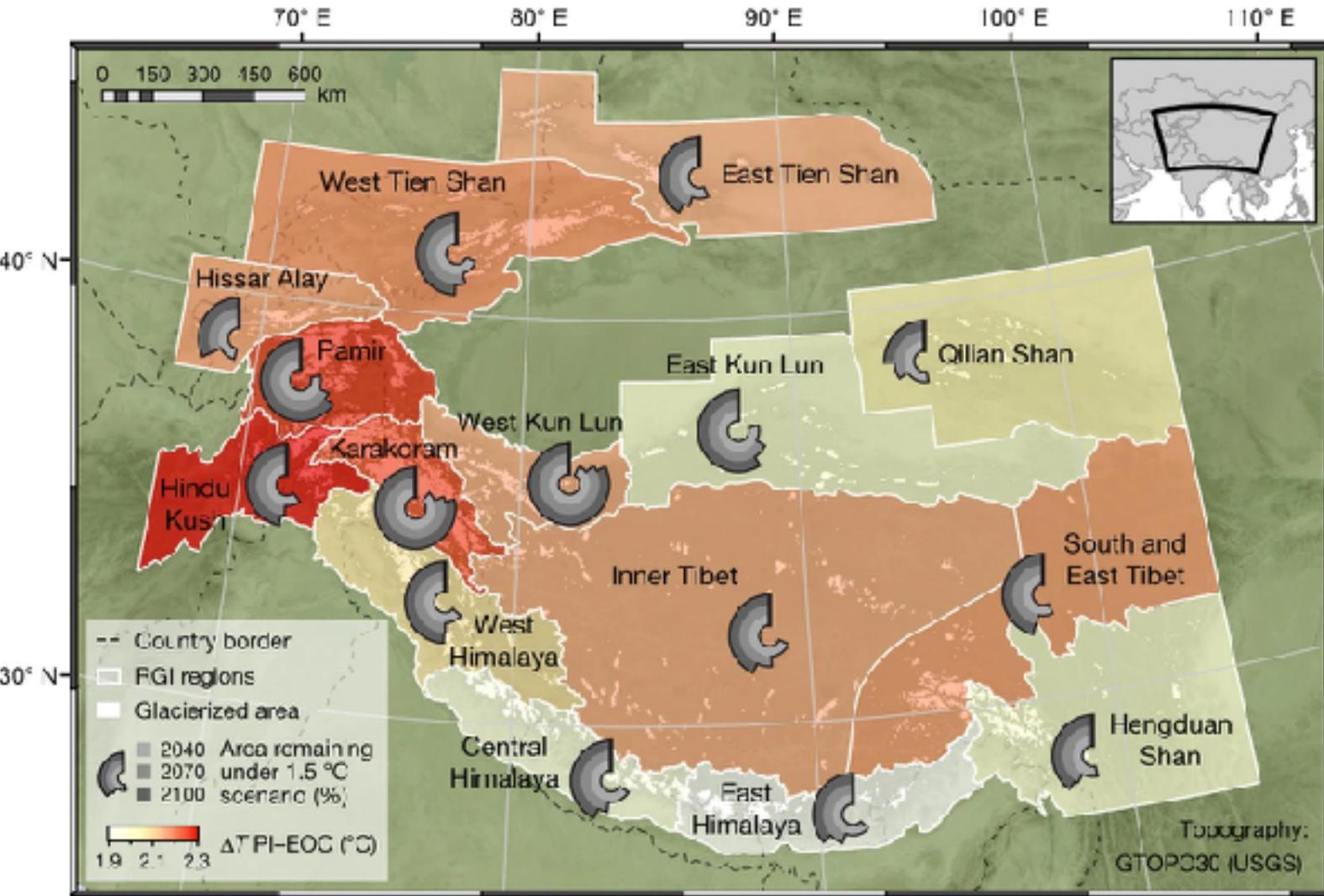
Hard to comprehend the scale of the flood disaster in Pakistan, the 5th most populated nation in the world.

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

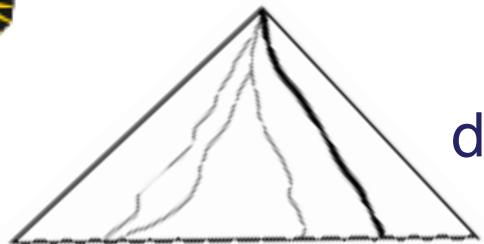
1/3 of the country is underwater.



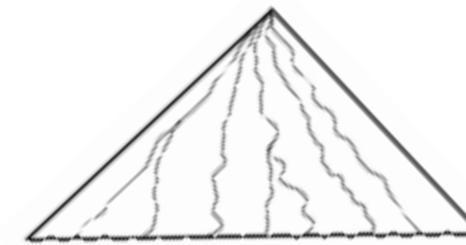
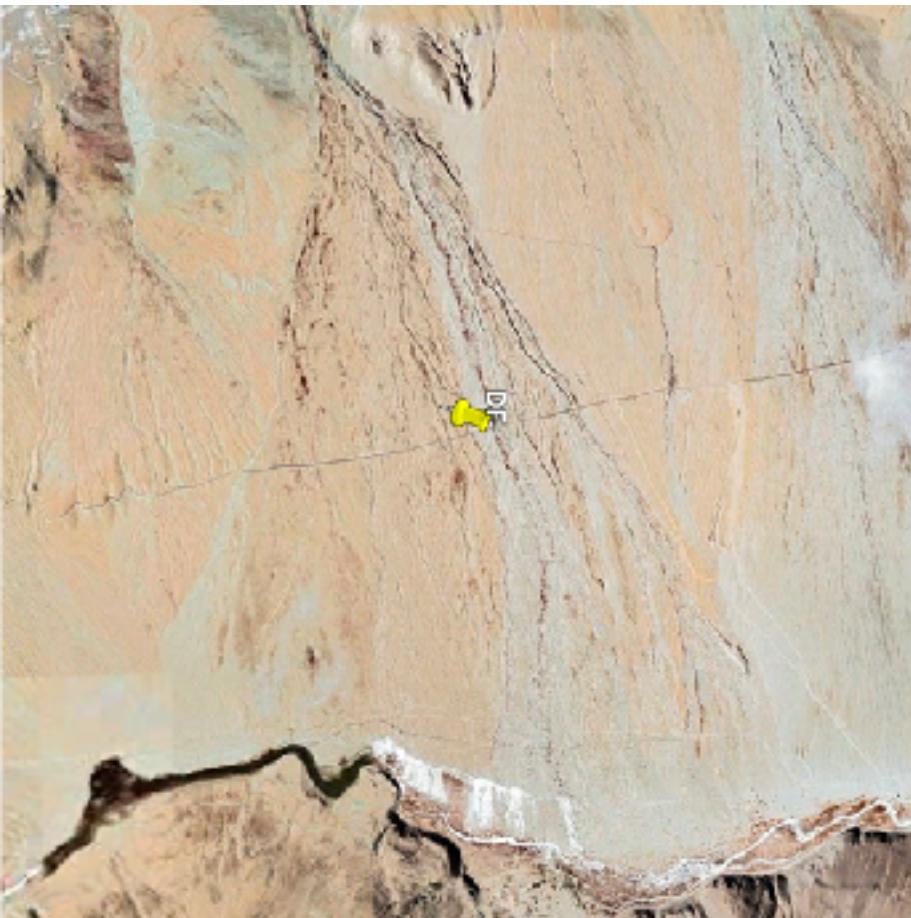
# What is happening in HMA?



*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

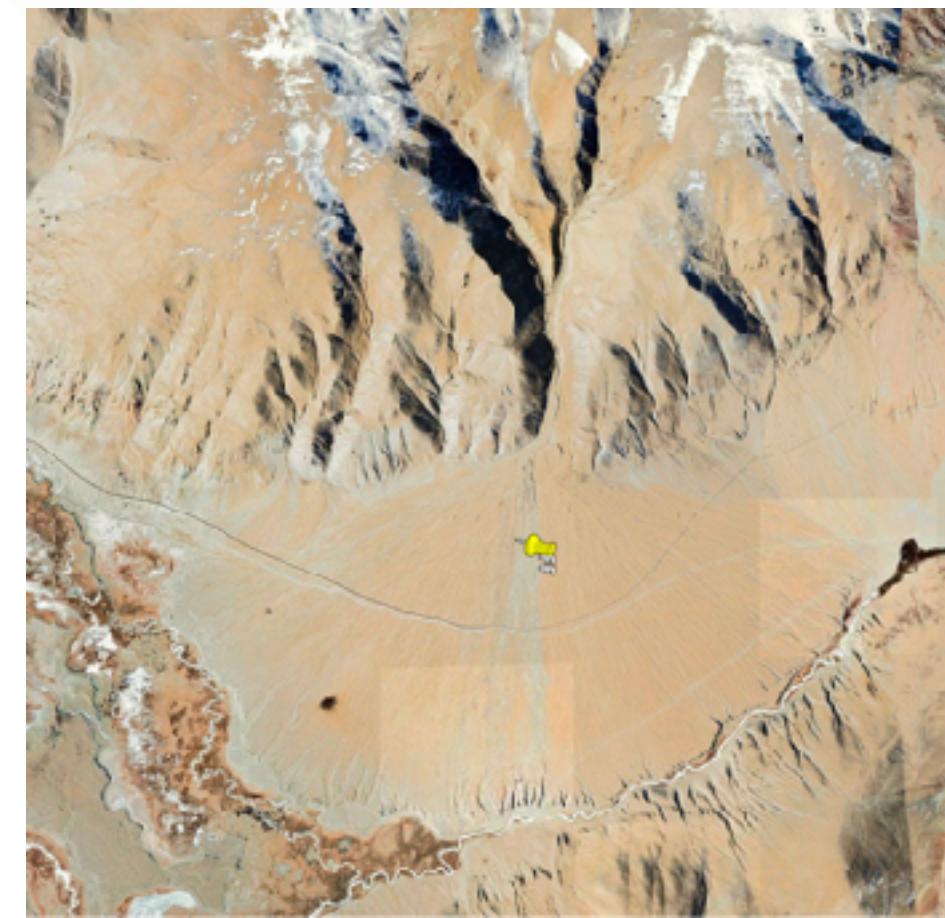


Image source: Google Earth



# 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



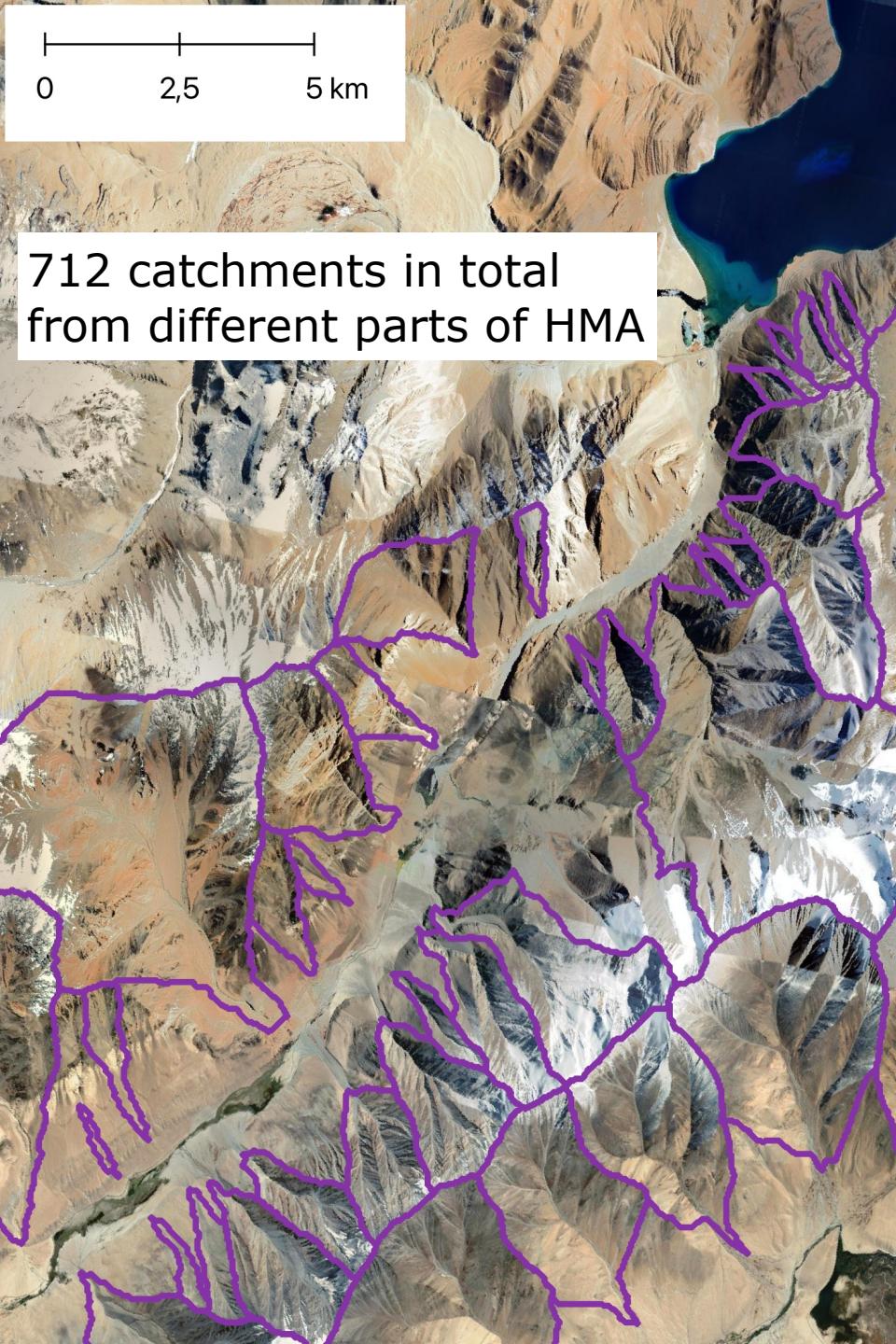
tune the  
model



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<sup>0.5</sup>)
- 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

...

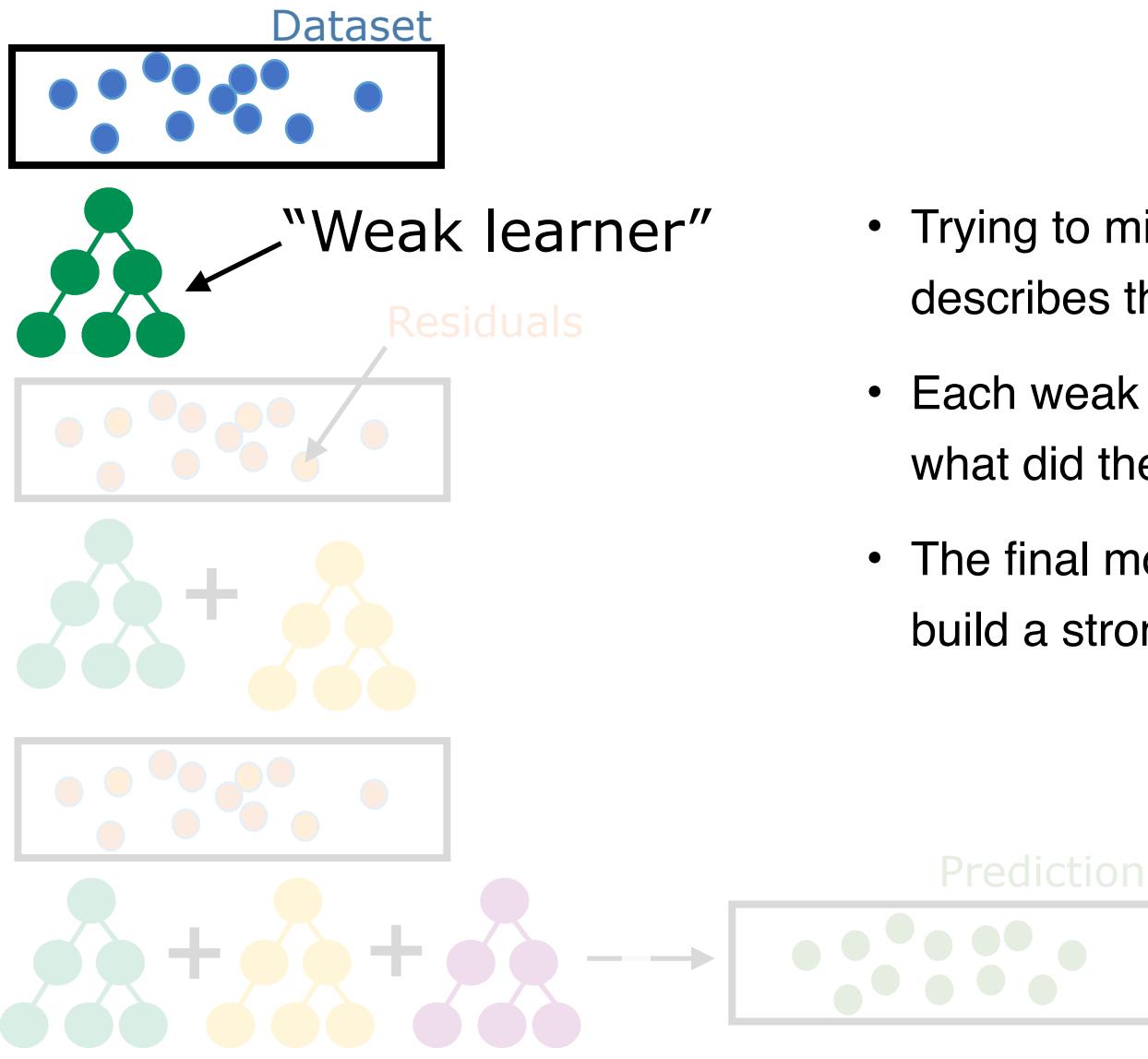
For thousands of data points and moderate dimensionality (99% of cases), gradient-boosted trees provide the necessary regression model  
[scikit-learn.org/stable/modules...](https://scikit-learn.org/stable/modules...)

They are robust to data distribution and support missing values (even outside MAR<sup>\*</sup> settings  
[arxiv.org/abs/1902.06931](https://arxiv.org/abs/1902.06931))

\* MAR = Missing At Random



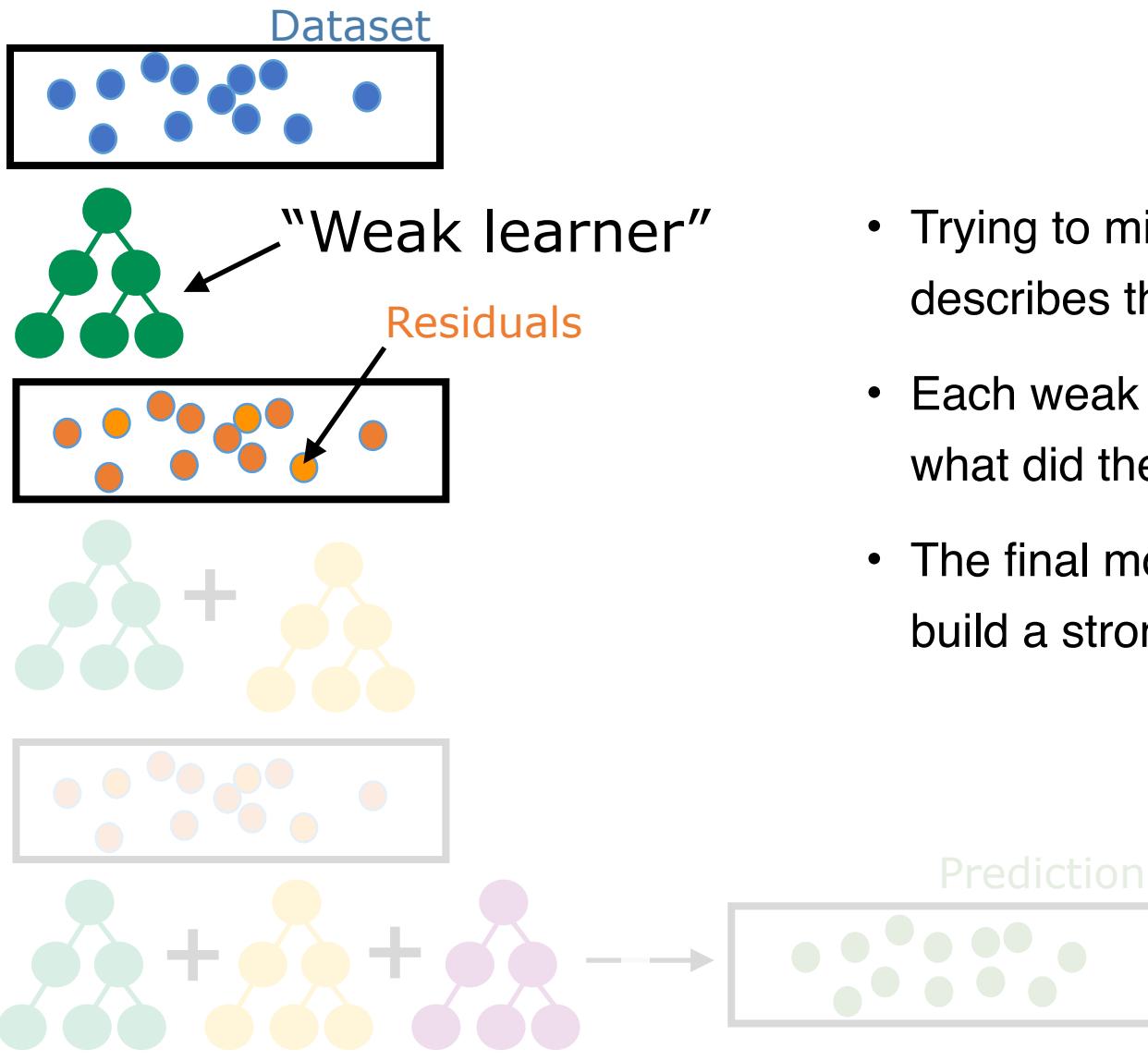
# Gradient boosted decision trees



- 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



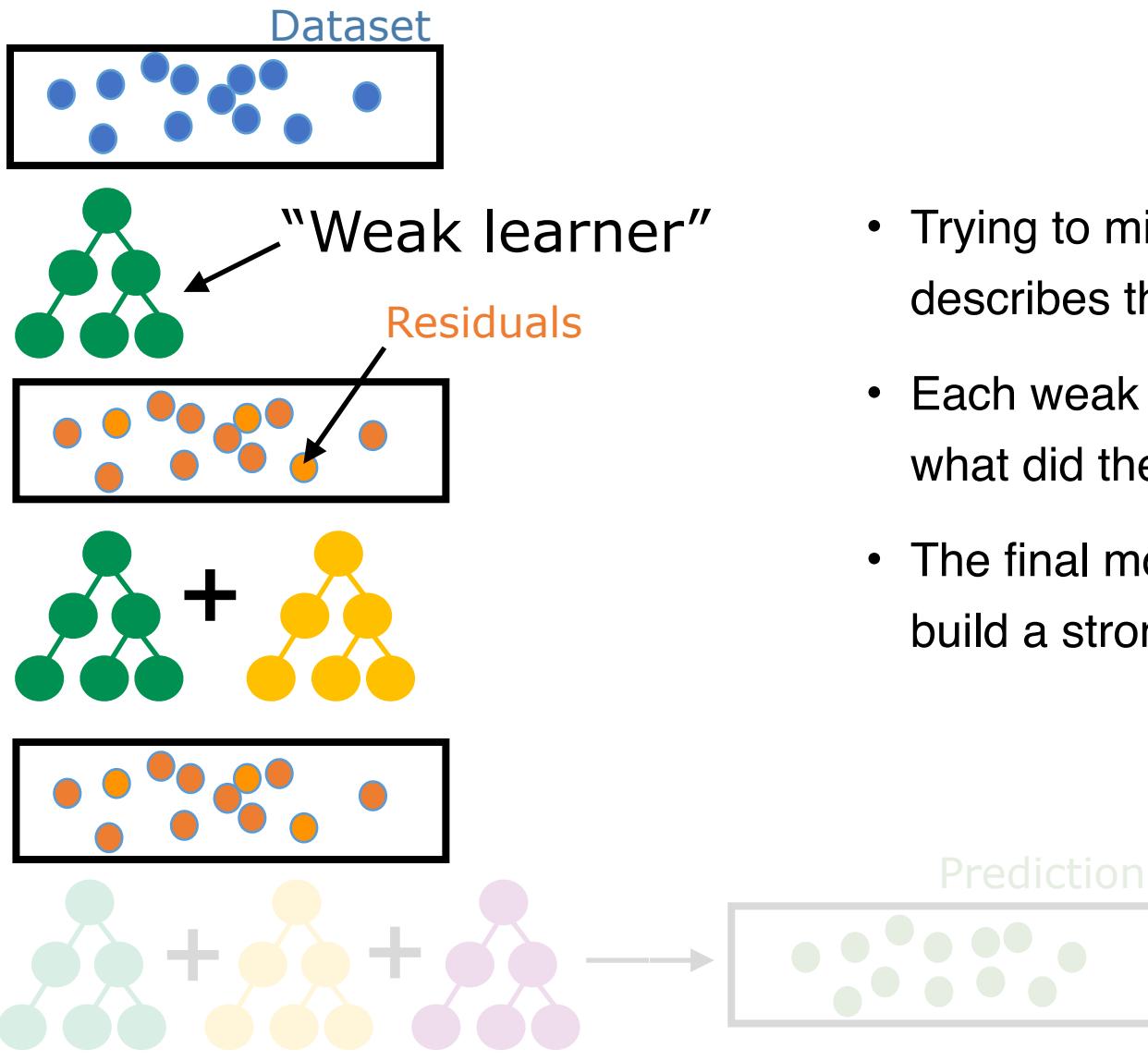
# Gradient boosted decision trees



- 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



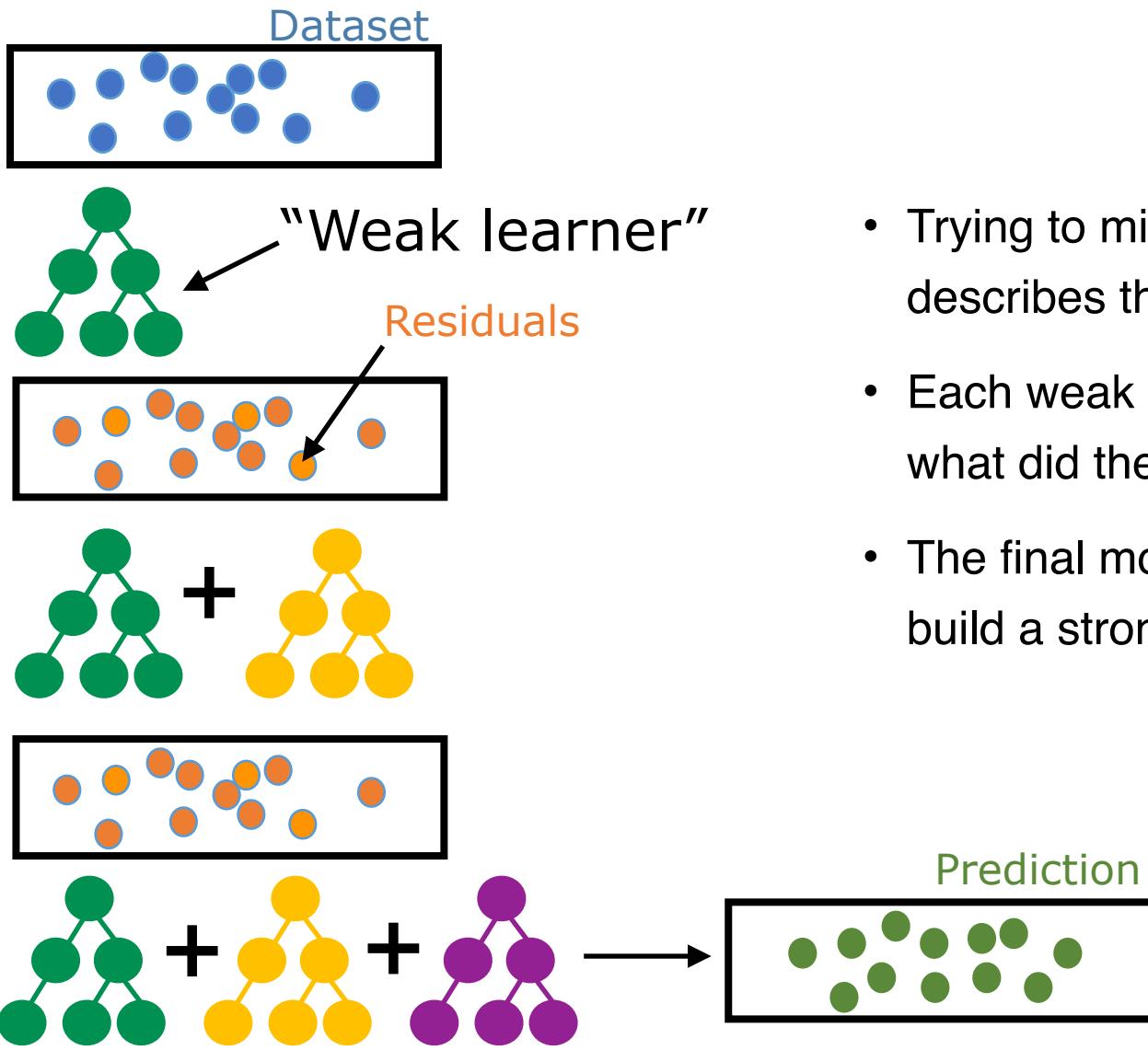
# Gradient boosted decision trees



- 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



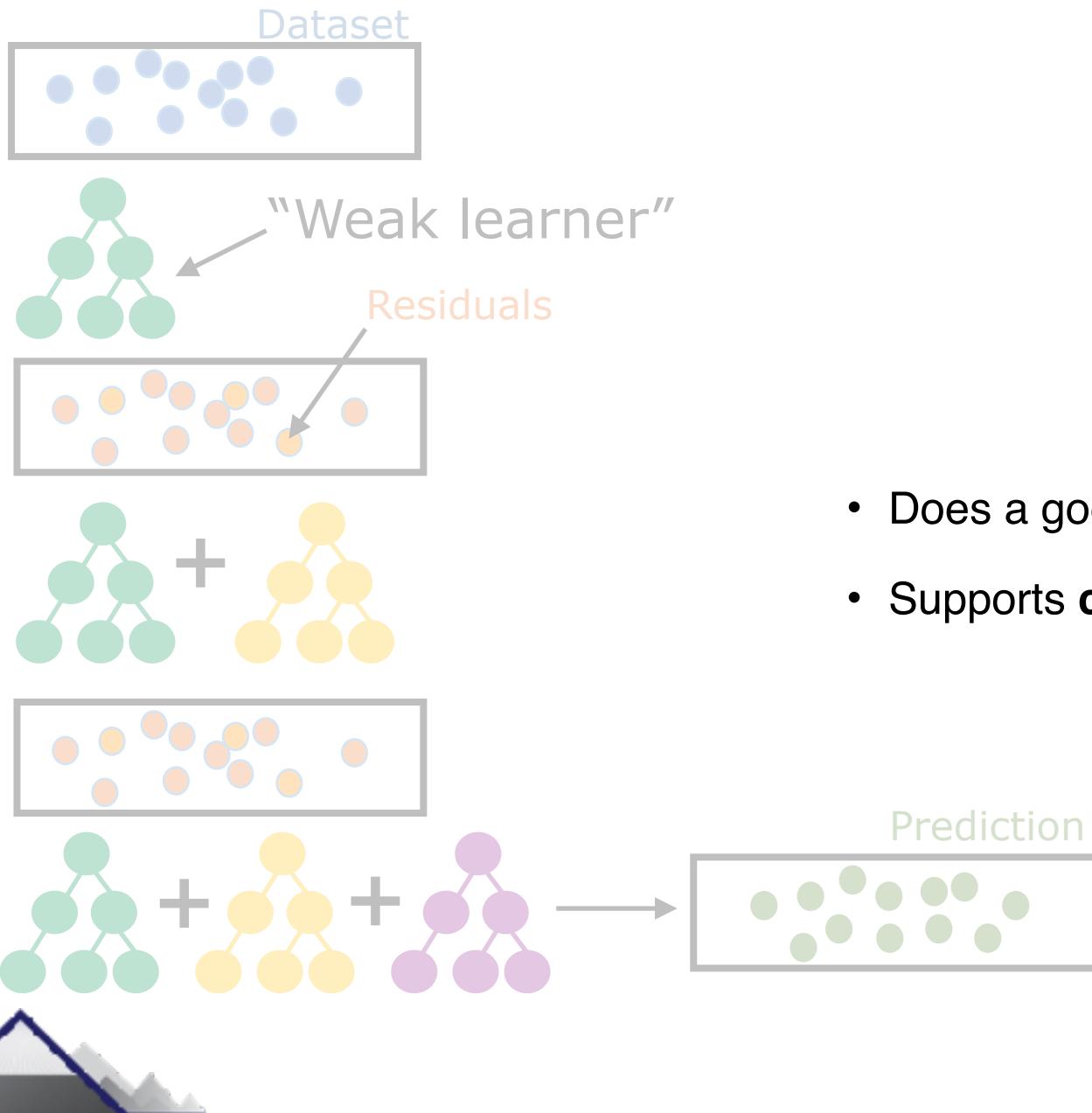
# Gradient boosted decision trees



- 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?

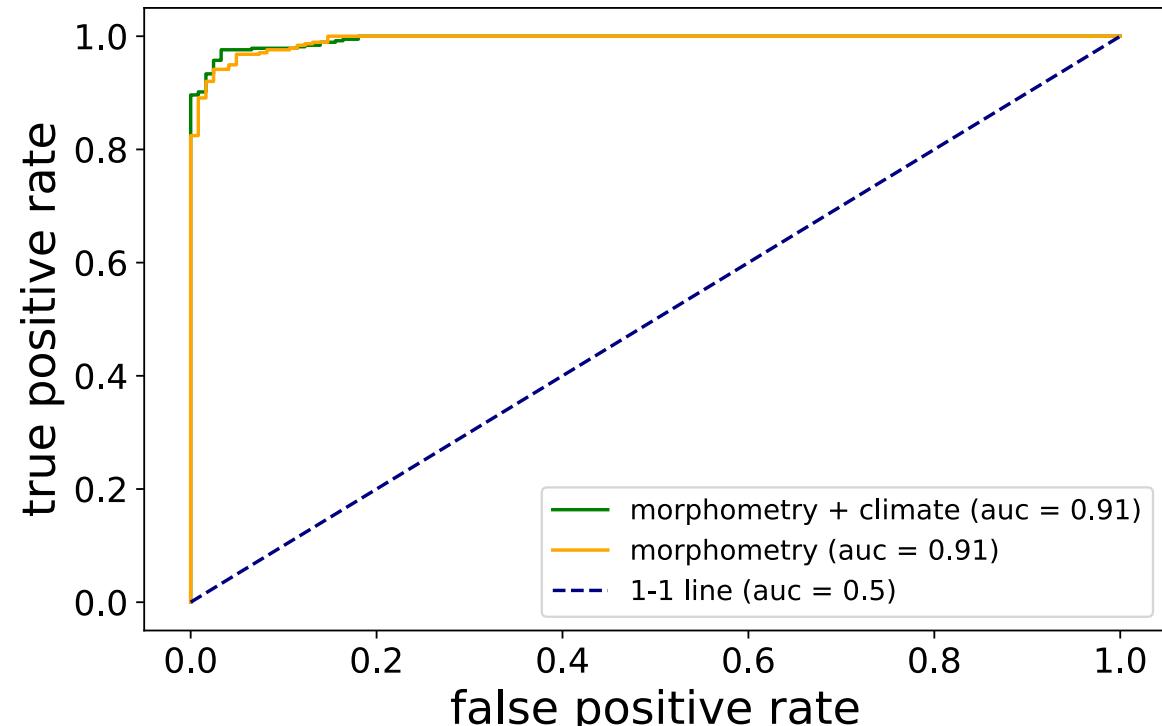
## Morphometric

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

## Morphometric + climate

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

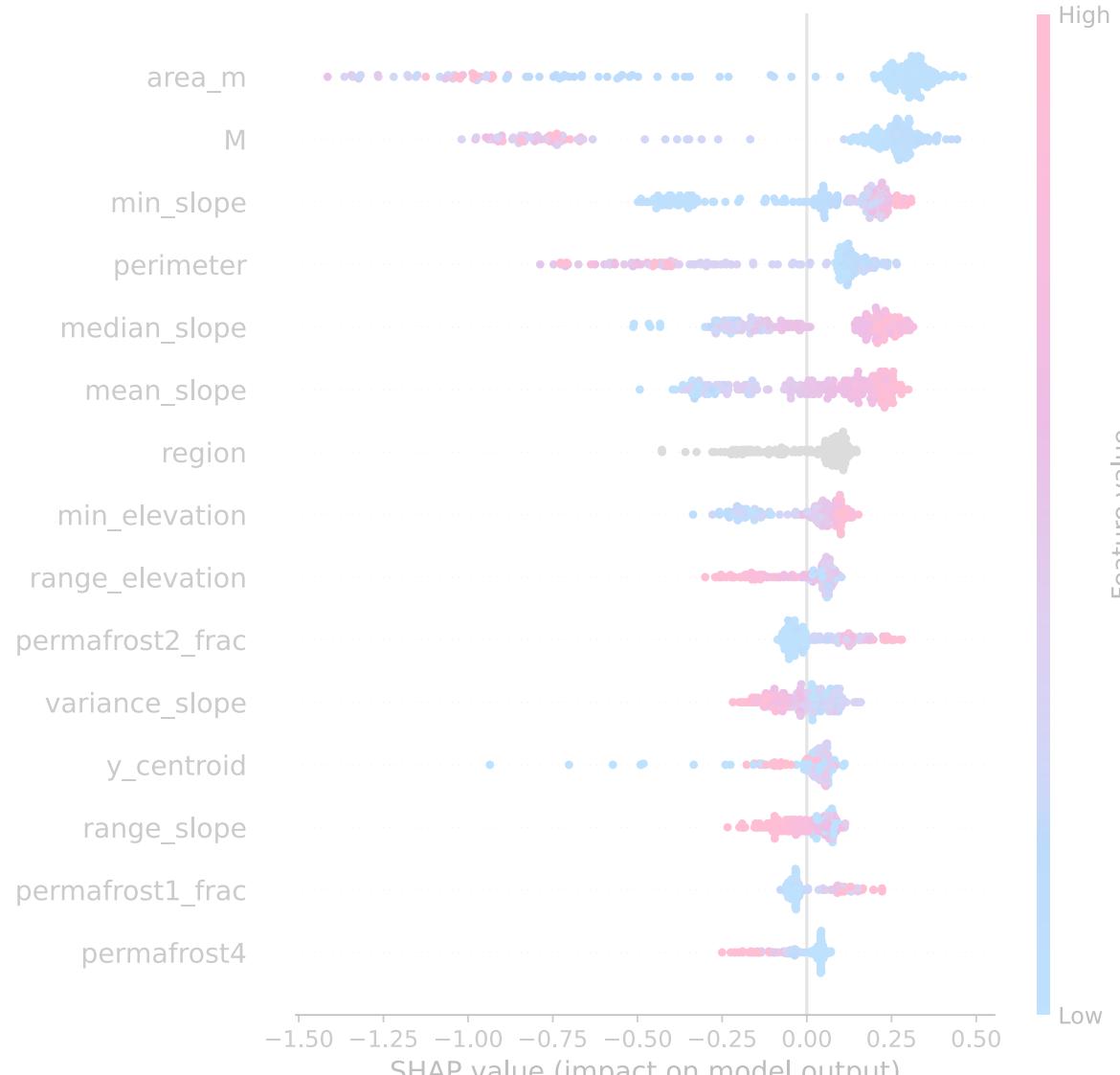
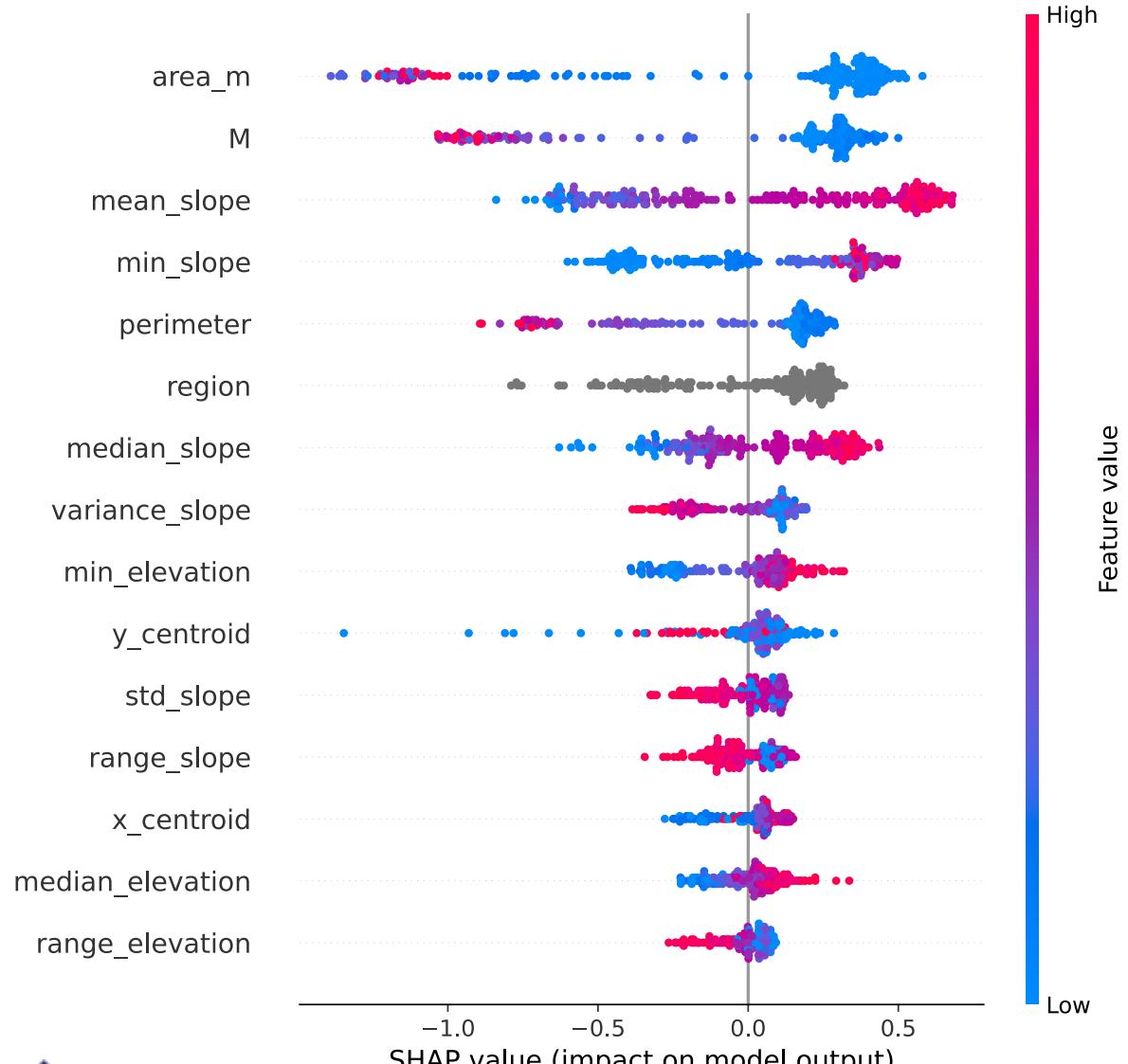
|    |    |
|----|----|
| TP | FN |
| FP | TN |



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

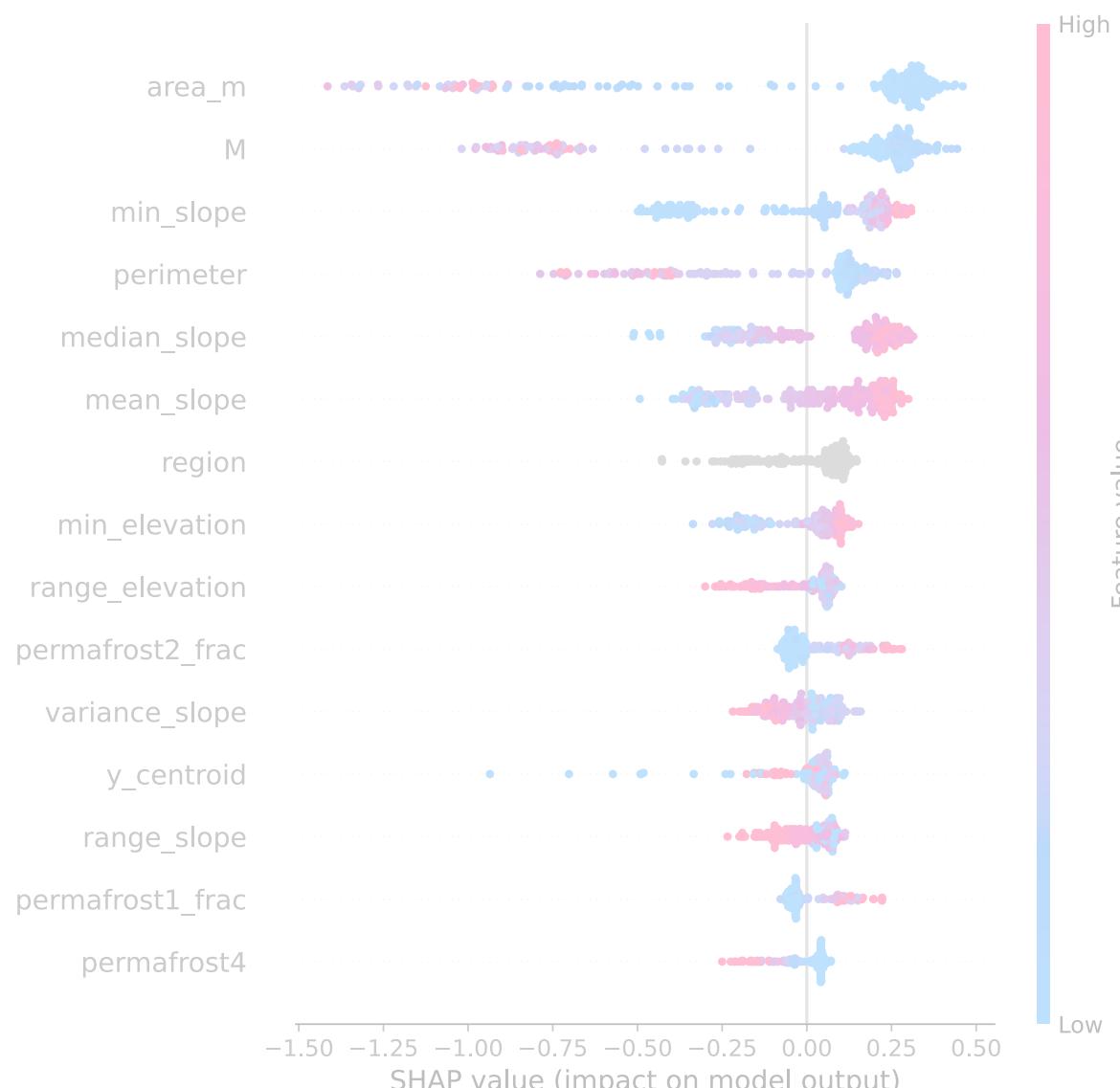
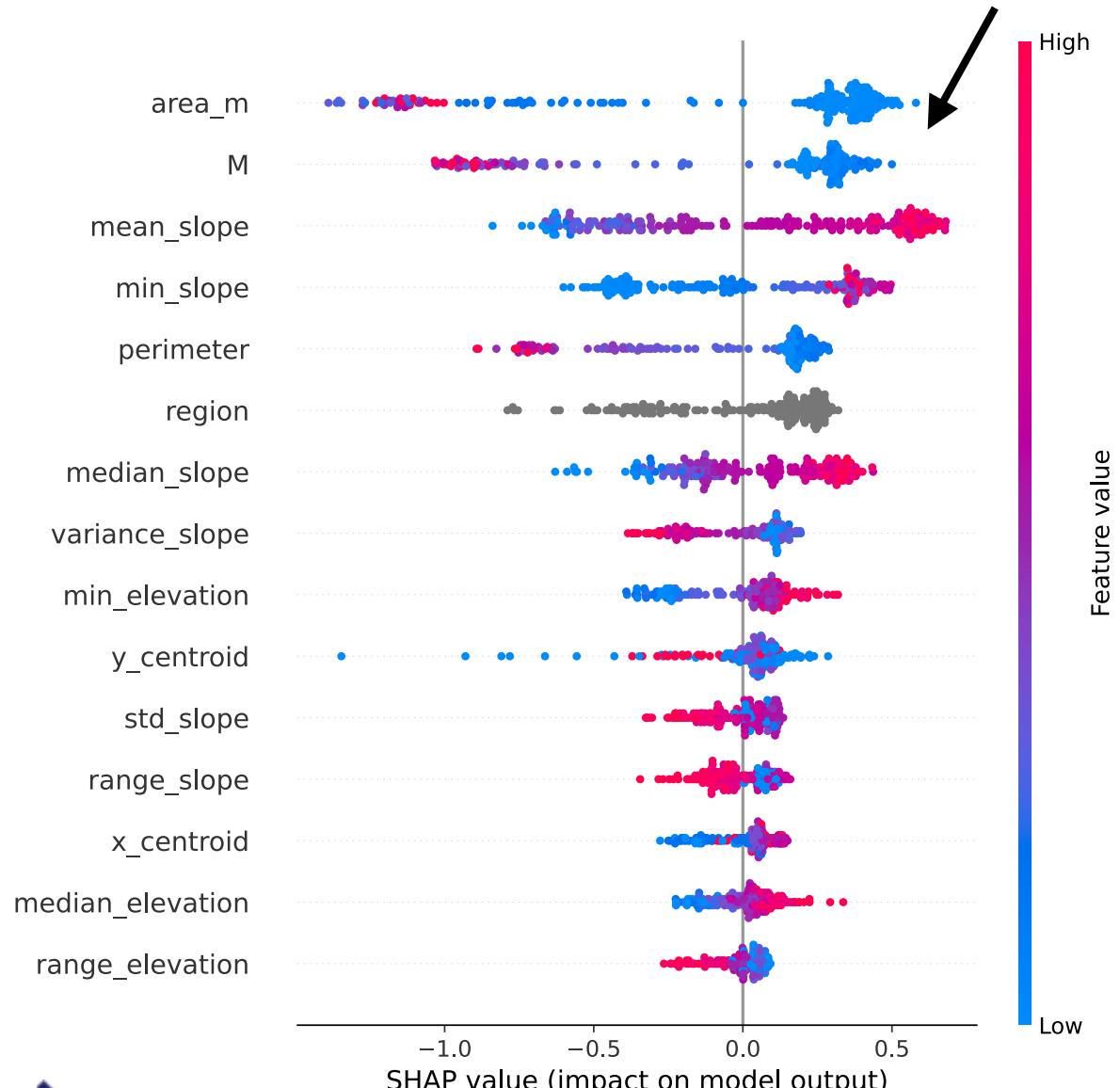


# Why does Catboost model make this predictions?



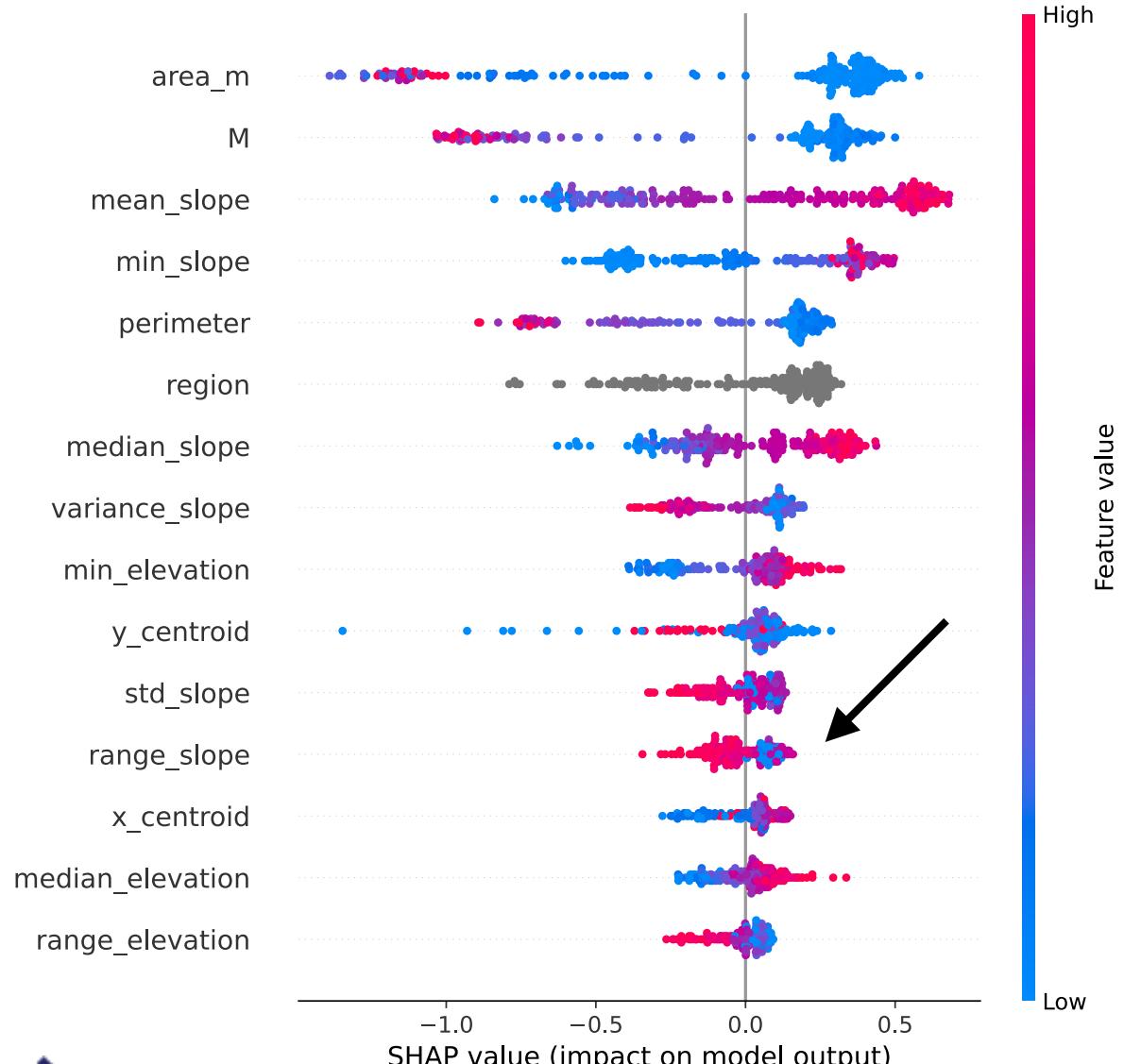


# Why does Catboost model make this predictions?

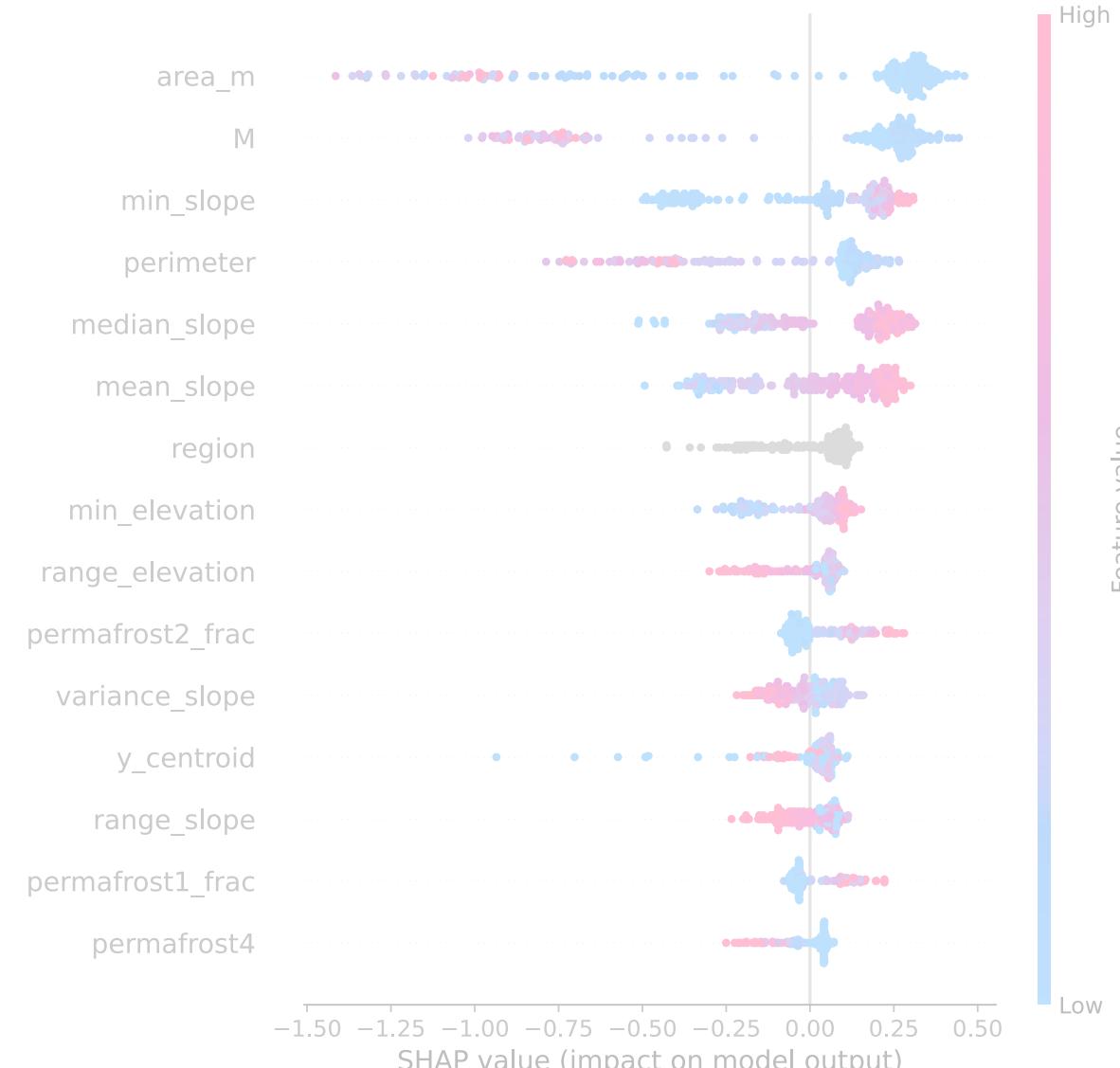




# Why does Catboost model make this predictions?



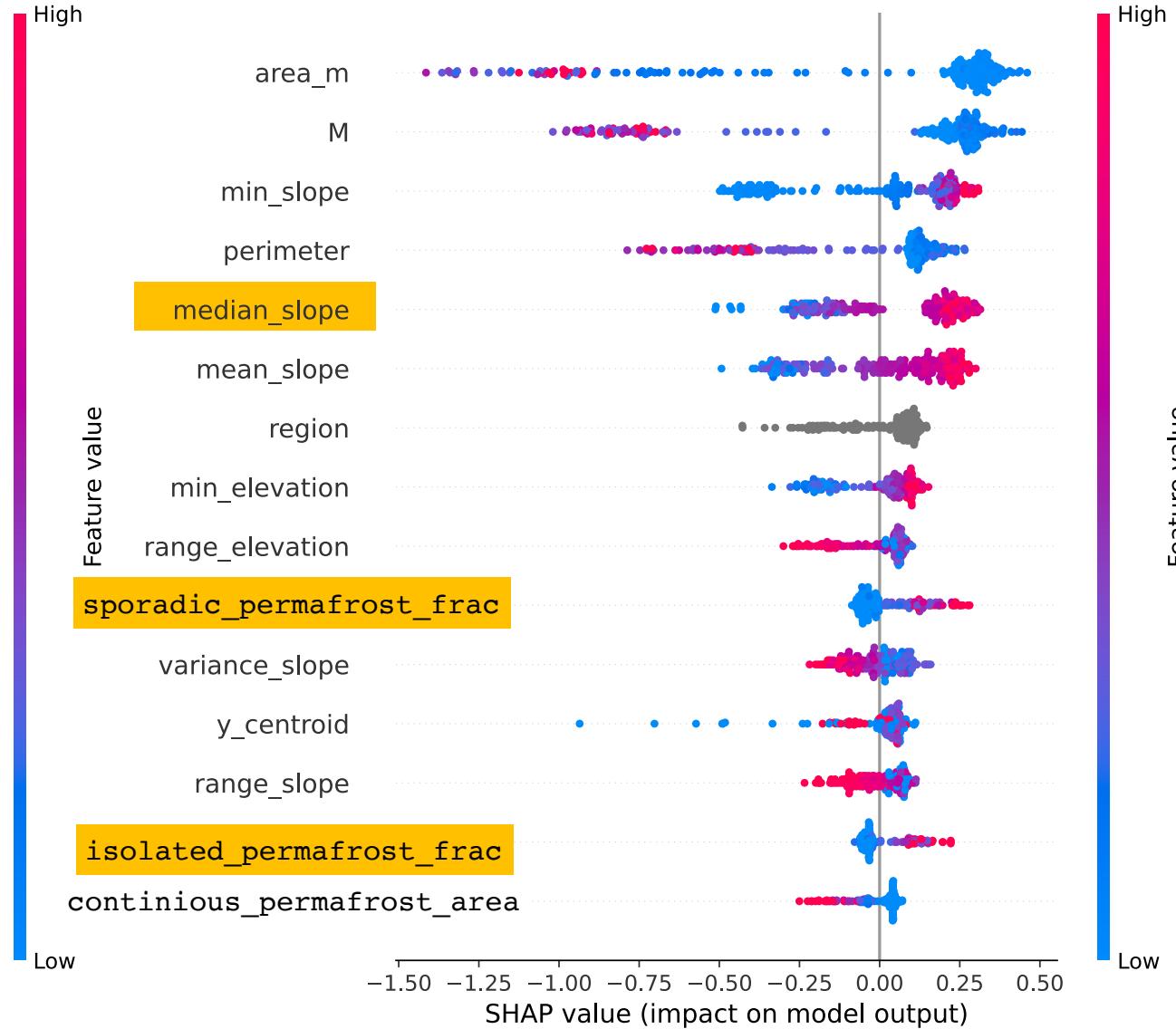
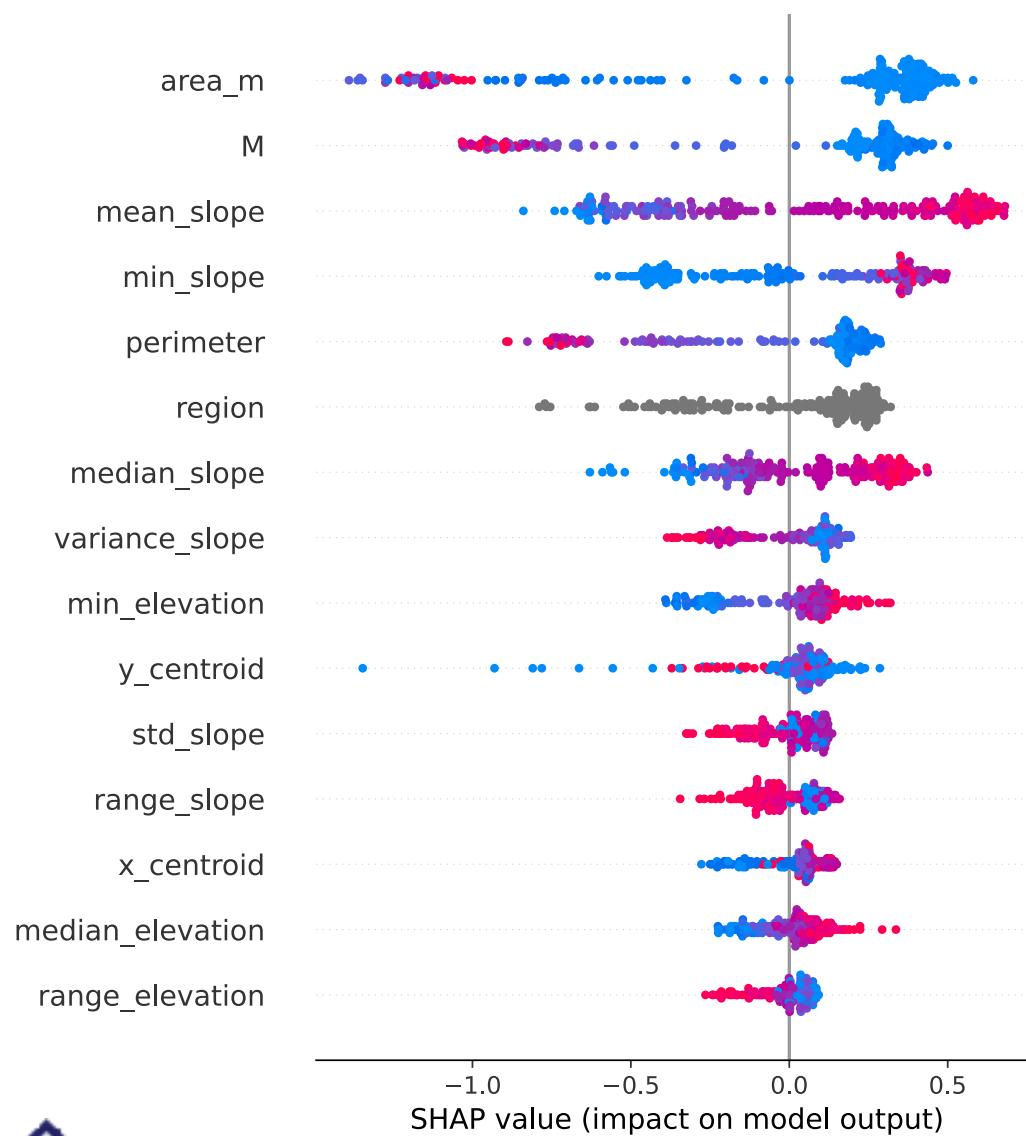
Morphometric



Morphometric + climate



# Why does Catboost model make this predictions?



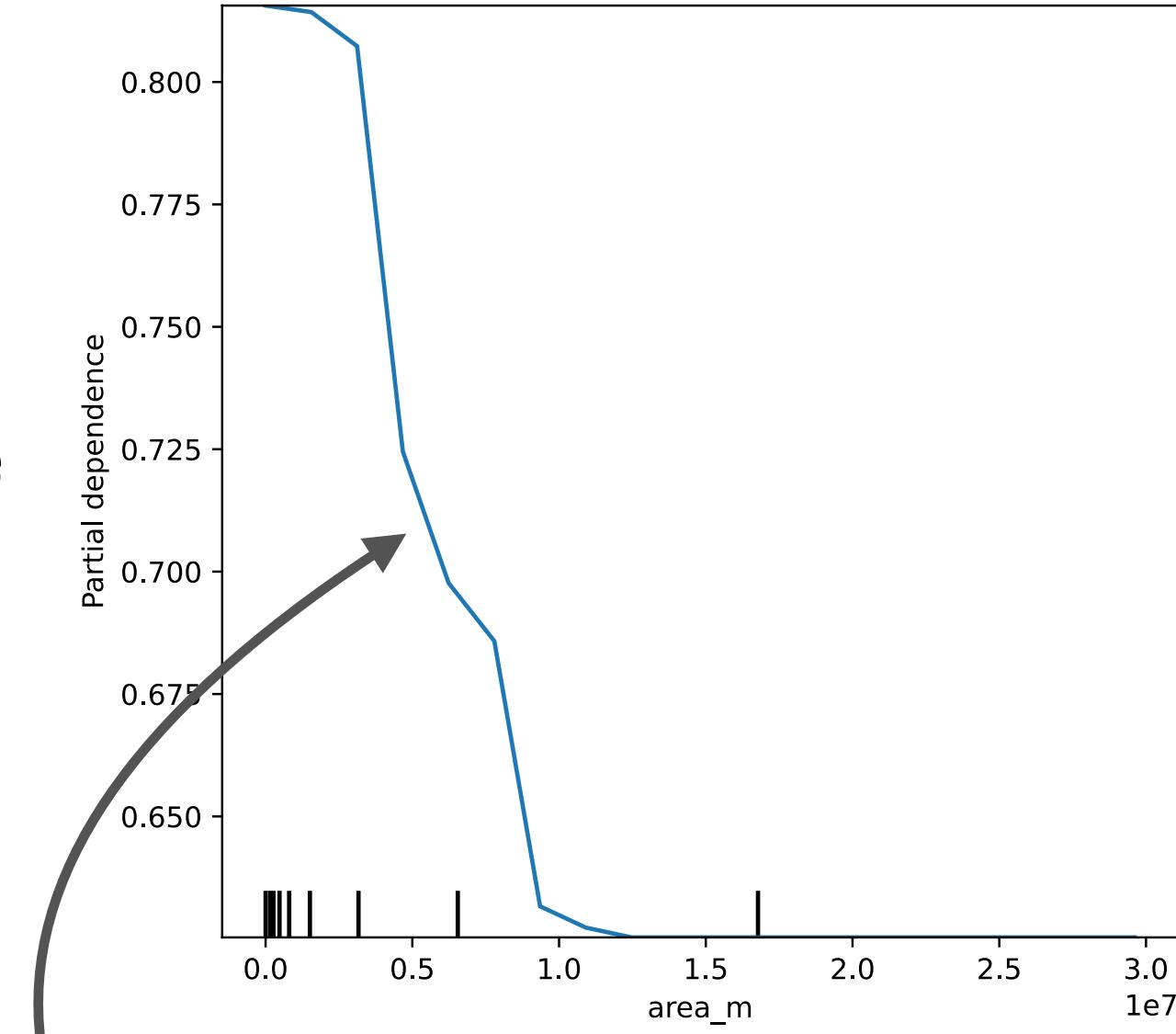


## What should we do to the catchment for it to “switch”:

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



