Automated detection of gravitational instabilities by combining seismology, satellite data and machine learning - example over the European Alps.

Charlotte Groult, Clément Hibert et Jean-Philippe Malet

contact : charlotte.groult@unistra.fr

Institut Terre & Environnement de Strasbourg | ITES | UMR 7063 de l'Université de Strasbourg & cnrs & ENGEES

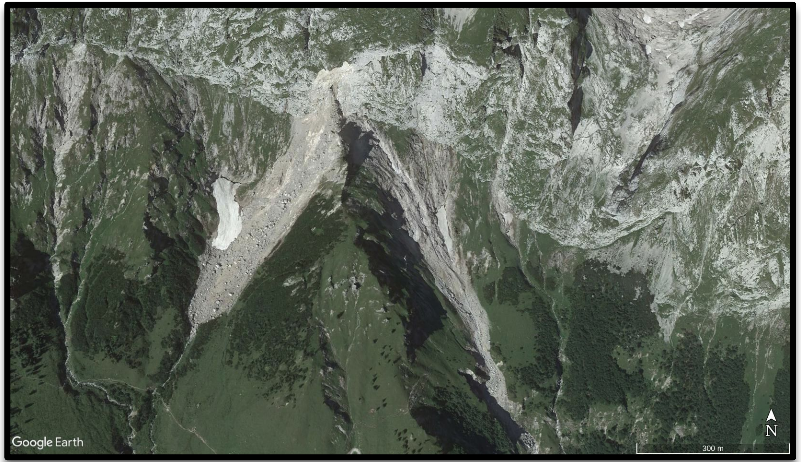
PROJET FINANCÉ PAR L'ANR · ANR · PROJECT FUNDED BY THE ANR

Photo Julien Chapuy

Introduction

❖ Context

- Climate change : permafrost melt
  - Number and intensity of landslides increasing in high latitude/altitude area
- Need to document landslide activity (detection, localization, link with environmental factors)
- manage risks in mountain and ecosystems



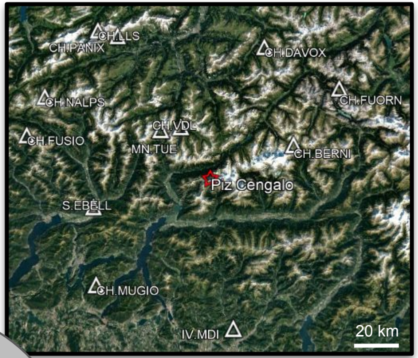
Event from ZAMG bulletin that occurred in Eck, Austria on 2018-05-07 4h18

❖ Objectives

- Develop an automated method to detect, classify and localize landslides
  - combine seismic and satellite data with a machine learning approach
- Build a catalog of landslide activity in the past 20 years
- Use the catalog - Explore correlations between landslide occurrences and environmental factors (weather, geology, slope…)

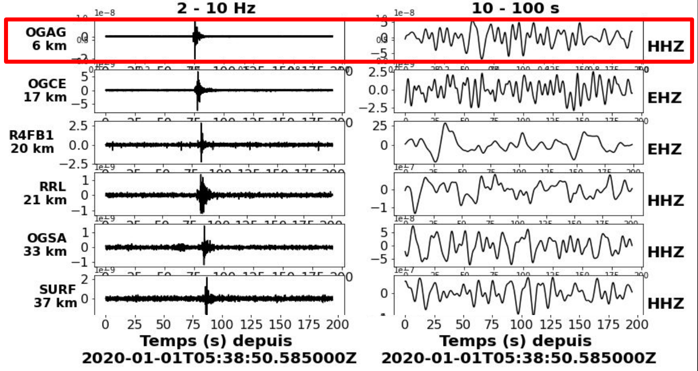
Seismic data for a well documented instrumental catalog

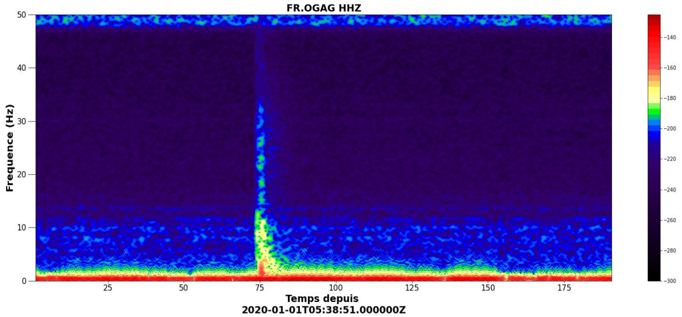
- Large regional scale - European Alps
- Data over long period (20 years)
- Exact date of occurrence



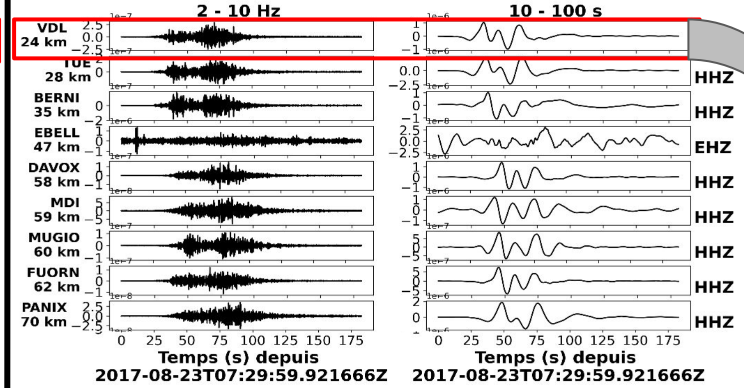
Map of an event and stations which record it

Earthquake

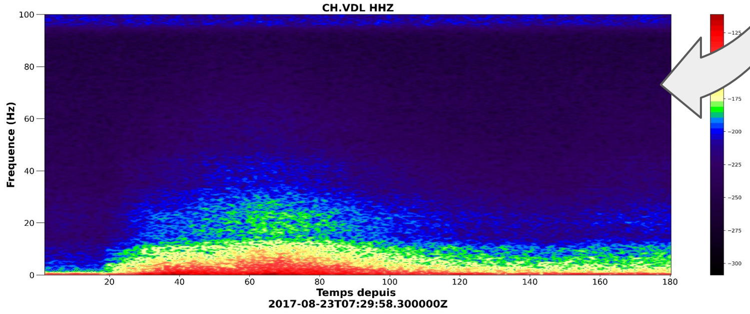




Landslide



From time-domain to frequency-domain



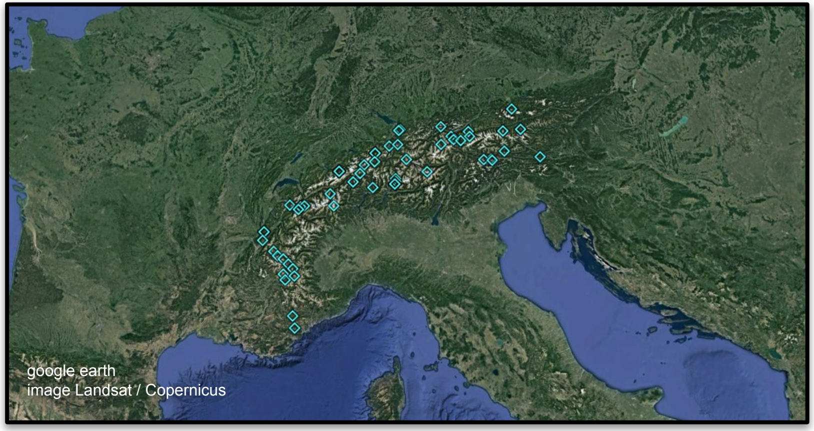
Spectrogram : Time-frequency representation

Database of known events that occurred in the Alps in the past 20 years

↳ to train and test my methods

❖ Landslides

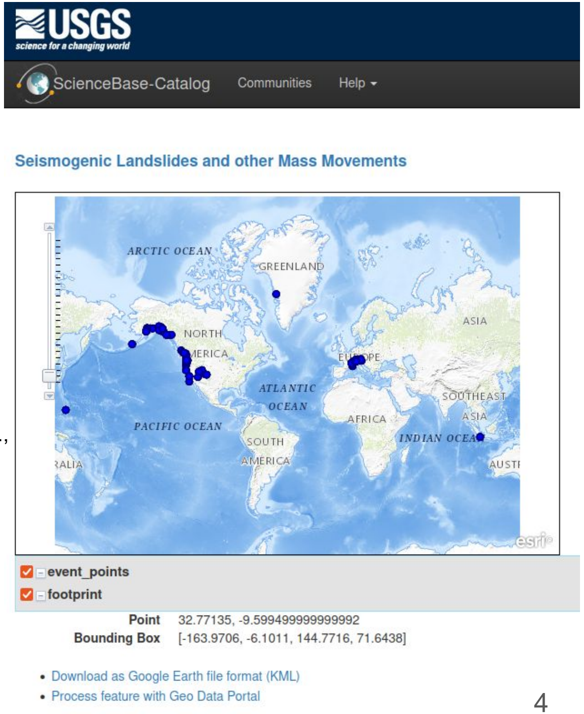
2589 seismic signals
generated by **68 events**
recorded by **547 stations**
since June 2002
sources : papers, media, ZAMG, Renass



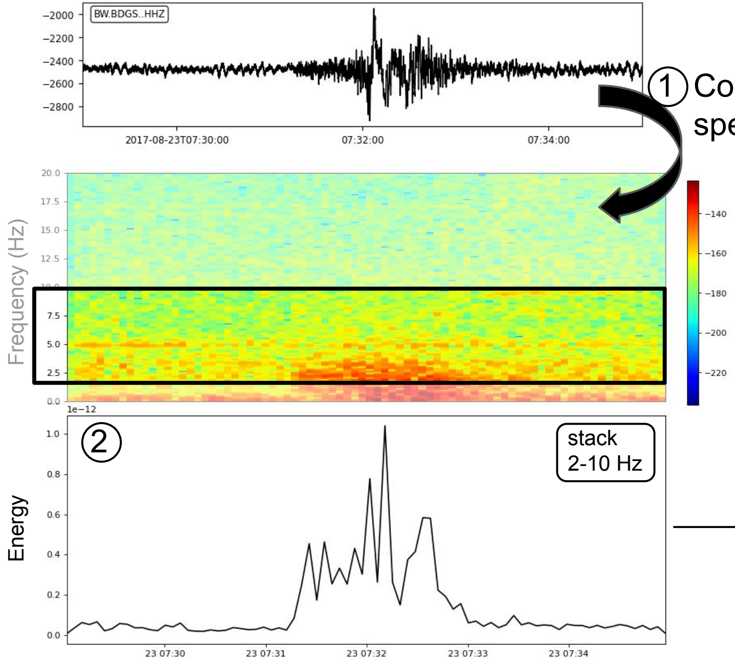
map of known seismogenic landslides in the Alps since 2002

Collaboration with USGS
Public database : events recorded by seismic network in North America, the European Alps and Pacific Islands

Collins, E.A., Allstadt, K.E., Groult, C., Hibert, C., Malet, J.-P., Toney, L.D., and Bessette-Kirton, E.K., 2022, Seismogenic Landslides and other Mass Movements: U.S. Geological Survey data release, https://doi.org/10.5066/P90VGCSK.



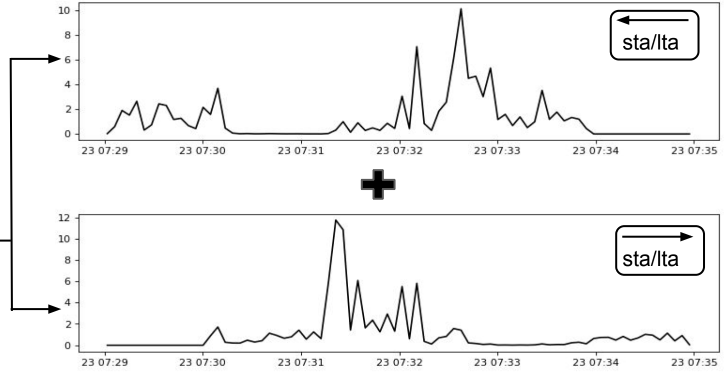
Detection : Spectral method



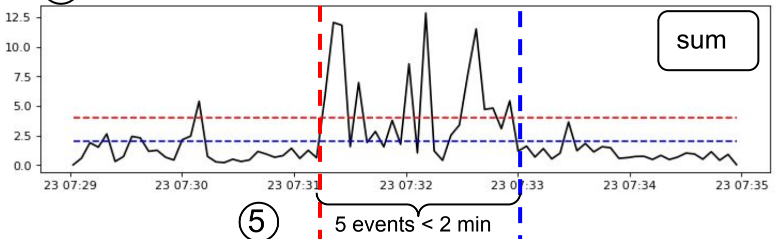
① Compute spectrogram

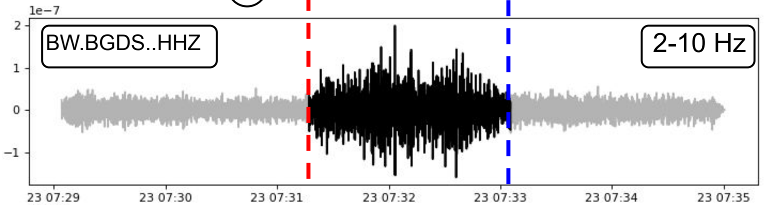
② Sum energy between 2 and 10 Hz

③ Compute sta/lta in two way



④ Detection on the sum of the two sta/lta



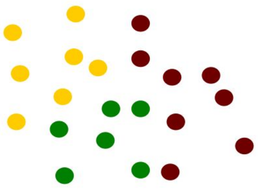


⑤ Gather detection with a gap less than 2 min

Classification : Random forest

❖ Compute model from the training set

Training set

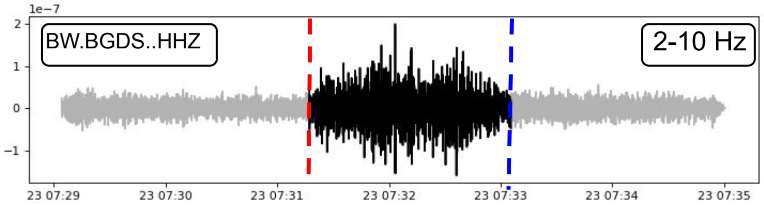


Features
(ex: length, waveform, …)

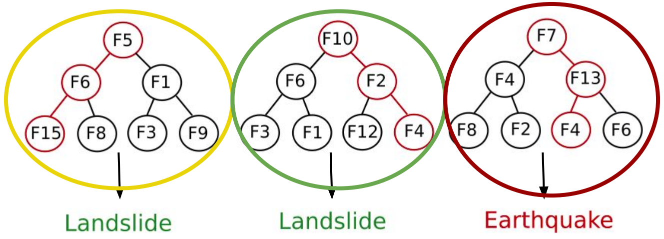
F1 F9
F2 F10
F3 F11
F4 F12
F5 F13
F6 F14
F7 F15
F8

- Compute N decision trees
- Each tree is computed from a subset of the training set and randomly selected features at each nodes

❖ Classify a new sample



computing features (F1, F2, … Fx)

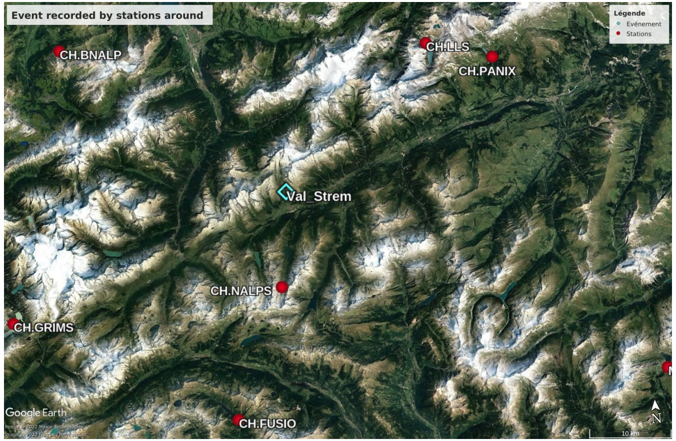


Compare new features to the model

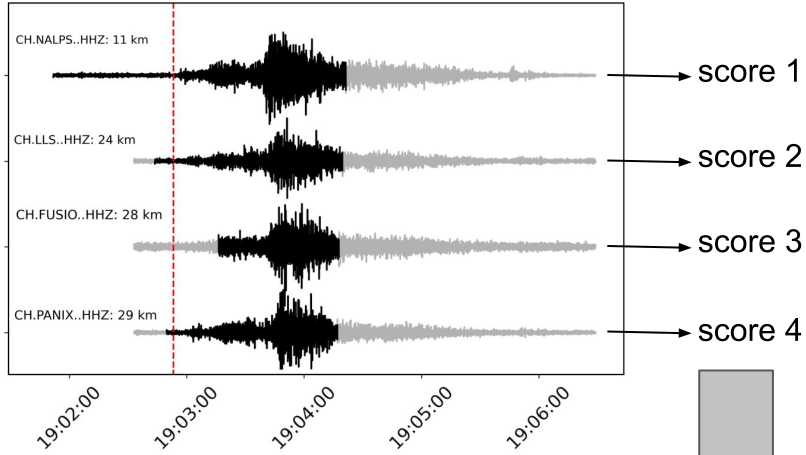
class = landslide
score landslide = 0.66

class and score are given by the majority of votes

Classification : Group signals by event and play with scores



One event generates signals on many stations around

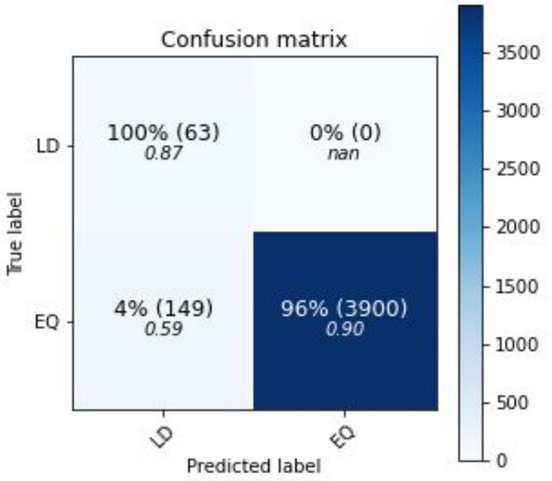


steps / processing chain	delete signals with low max/med and max/mean	detections ≥ 3 stations	score min	score max	median score	majority of votes	classification ≥ 2 stations	points (≥ 10)
1	1	2		3			4	
2	1		2			4		
3	1	2	3			4		
4	1	2	3	4				
5	1	2		3 (10 pts)	4 (7 pts)	5 (3 pts)		✓
6	1	2		4 (7 pts)	5 (3 pts)	3 (10 pts)		✓
7	1	2		5 (3 pts)	3 (10 pts)	4 (7 pts)		✓

- Events are classified by considering all signals that refer to it
- Many possibilities are investigated

❖ Test of random forest classifier with the training set

- 50 % signals from the training set is used to build the model (forest of decision trees) - the rest is used to test the classifier
- Delete events with mean score < 0.6



- Accuracy = 96%
- Precision :
  LD = 30%
  EQ = 100%
- Recall :
  LD = 100%
  EQ = 96%
- Score F1:
  LD = 46
  EQ = 98

❖ Test of random forest classifier on continuous data

- Run detection and classification on 34 days which include known events

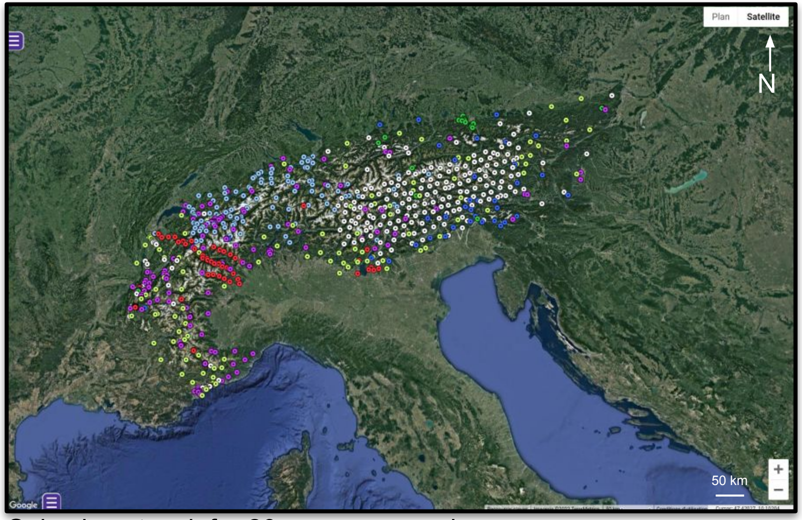
Processing chain	Number of known events found / number of total events tested	Number of gravity events found on the 34 days tested
1	18 / 34	1333
2	13 / 34	262
3	18 / 34	1118
4	17 / 34	1211
5	22 / 34	2030
6	23 / 34	2634
7	22 / 34	2168

❖ Conclusion

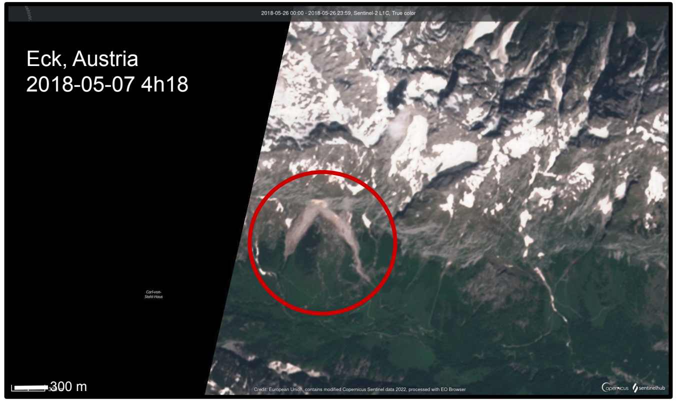
- Tests on training set provide a good identification rate (100 % for LD - 96% for EQ)
- Results on continuous data need to be improved : Find a processing chain which properly classify all the known events and limit false alarms

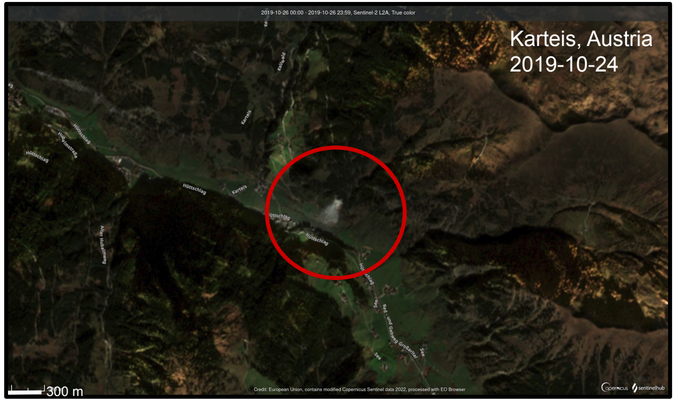
❖ Perspectives

- Include satellite data : constrain localization and volume
- Apply the method on 20 years on the whole network
- Explore correlations between landslide occurrences and environmental factors (weather, geology, slope…)



Seismic network for 20 years processing





Sentinel-2 images for satellite processing