Self-supervised learning for (environmental) seismology

1ère Journées Epos France – Saint-Jean-Cap-Ferrat

10/11/2023

Clément Hibert | hibert@unistra.fr

Joachim Rimpot, Jean-Philippe Malet, Germain Foretier, Jonathan Weber, Lise Retailleau, Jean-Marie Saurel, Antoine Turquet, Tord Stangeland

et al.

École et observatoire des sciences de la Terre | Université de Strasbourg

IRIMAS Université Haute-Alsace
Institut de Recherche en Informatique, Mathématiques, Automatique et Signal

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

IPGP

OVPF

NORSAR

How can seismology help to understand environmental processes ?

- Detection and identification of active areas (*where? what?*)
- Monitoring to alert on possible risks (*when?*)
- Understanding the influence of different forcings (meteorological, climatic, tectonic) (*why?*)

INTRODUCTION | ENVIRONMENTAL SEISMOLOGY

How can seismology help to understand environmental processes ?

- Detection and identification of active areas (*where? what?*)
- Monitoring to alert on possible risks (*when?*)
- Understanding the influence of different forcings (meteorological, climatic, tectonic) (*why?*)

Detection & localisation of seismic sources :

- Global Scale : large events (landslides, calving events, etc.)
- Regional and local scale : rockfalls, lahars, debris flows, avalanches
- Endogeneous seismicity : landslides, glaciers, etc.

Characterization of the properties and dynamics of the sources :

- Inversion and modelisation with long period waves (>30-40 s)
- Statistical scaling laws with short period waves (<1 s)

How can seismology help to understand environmental processes ?

- **Detection** and **identification** of active areas (*where? what?*)
- **Monitoring** to alert on possible risks (*when?*)
- **Understanding** the influence of different forcings (meteorological, climatic, tectonic) (*why?*)

Detection & localisation of seismic sources :

- **Global Scale :** large events (landslides, calving events, etc.)
- **Regional and local scale :** rockfalls, lahars, debris flows, avalanches
- **Endogeneous seismicity** : landslides, glaciers, etc.

Characterization of the properties and dynamics of the sources :

- Inversion and modelisation with **long period waves** (>30-40 s)
- Statistical scaling laws with **short period waves** (<1 s)

DETECTION | CLASSIFICATION

Objective : Find rare events in continuous data

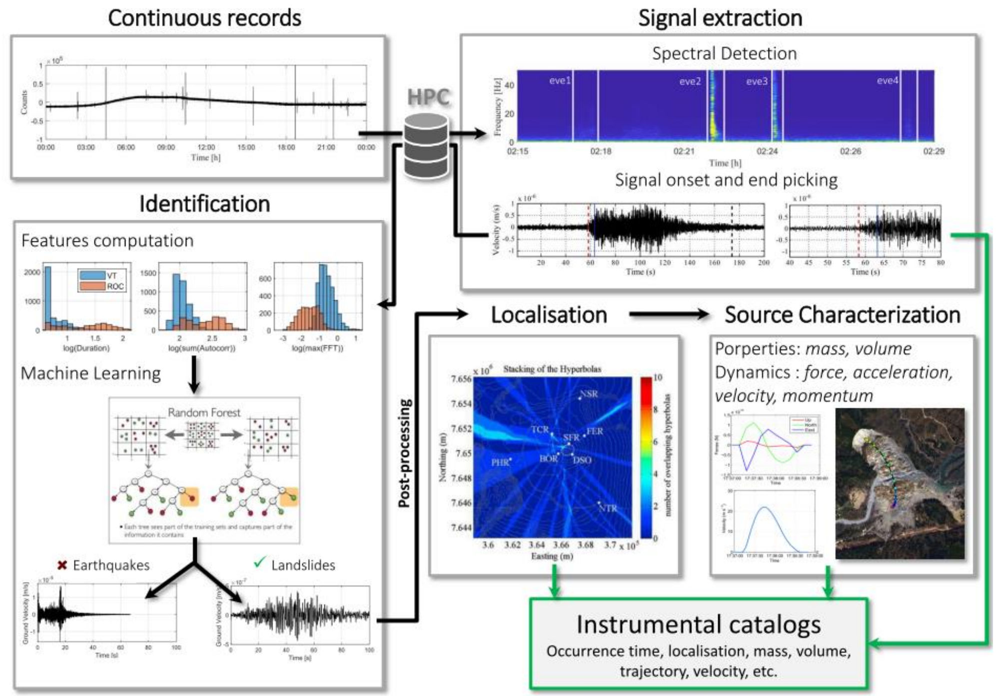
- Restrospectively
- In real-time

How to find rare events in continuous streams of data ?

Supervised classification :

Which algorithms ?

Which features ?



Many constraints :

- **Robust**, **versatile**, **portable** to different contexts and for different sources
- Able to be trained with **few examples**
- Able to produce a very high rate of good identification even with a **reduced network** (1 or 2 sensors, 1 component)
- Able to be efficient with sometimes very **unbalanced data sets**

Testing ensemble algorithms + curated features

Local scale :

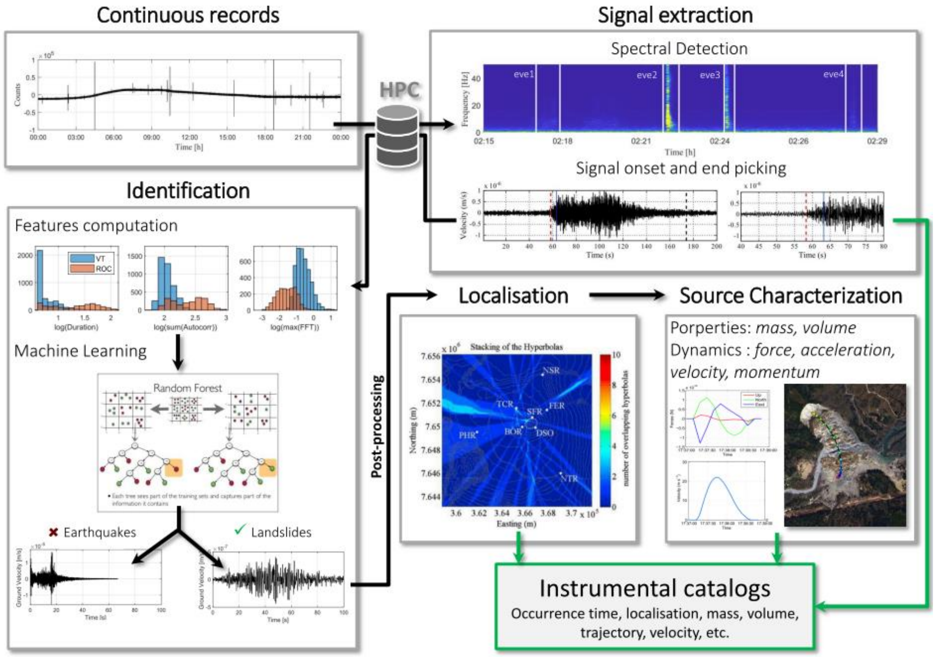
- Super-Sauze [Provost et al., 2017] – 4 classes, ~900 eve. Success rate : 90%
- Piton de la Fournaise volc. [Maggi et al., 2017; Hibert et al., 2017] – 2-8 classes, 13000+ eve. : 90-95+%
- La Clapière – 4 classes, ~11100 eve. : 92%
- Séchilienne – 4 classes, ~130000 eve. : 91%
- Knipovich Ridge [Domel et al., 2023] : 87%

Regional scale :

- Alaska [Hibert et al., 2019]
- Alps : WIP [Groult et al., in prep.] > ANR HighLand
- Greenland [Pirot et al., 2023]

Processing streams of data :

- Illgraben/Piz Cengalo [Wenner et al., 2021; Chmiel et al., 2021]: 80-90%
- DAS [Huynh et al., 2022; in prep.] : 87%
- Super-Sauze [Rimpot et al, in prep.]



CLASSIFICATION

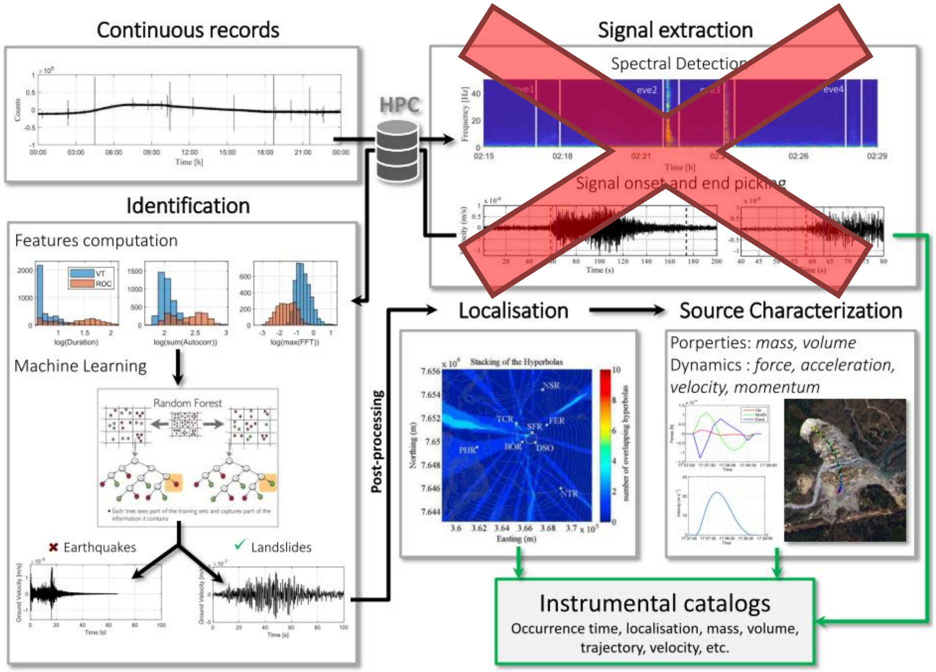
Testing ensemble algorithms + curated features

Local scale :

- Super-Sauze [Provost et al., 2017] – 4 classes, ~900 eve. Success rate : 90%
- Piton de la Fournaise volc. [Maggi et al., 2017; Hibert et al., 2017] – 2-8 classes, 13000+ eve. : 90-95+%
- La Clapière – 4 classes, ~11100 eve. : 92%
- Séchilienne – 4 classes, ~130000 eve. : 91%
- Knipovich Ridge [Domel et al., 2023] : 87%

Regional scale :

- Alaska [Hibert et al., 2019]
- Alps : WIP [Groult et al., in prep.] > ANR HighLand
- Greenland [Pirot et al., 2023]

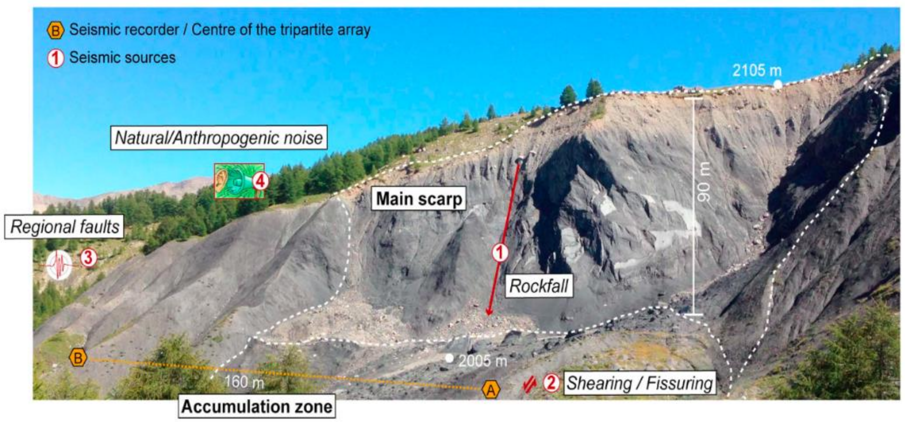


Processing streams of data :

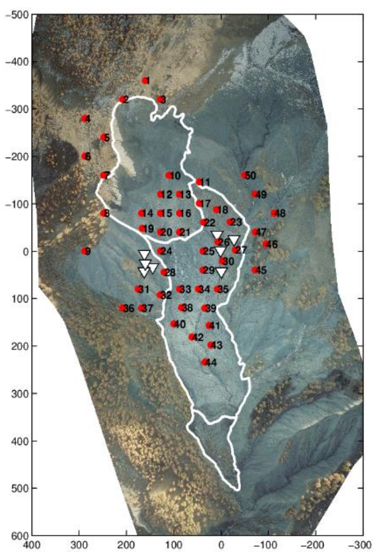
- Illgraben/Piz Cengalo [Wenner et al., 2021; Chmiel et al., 2021]: 80-90%
- DAS [Huynh et al., 2022; in prep.] : 87%
- Super-Sauze [Rimpot et al, in prep.]

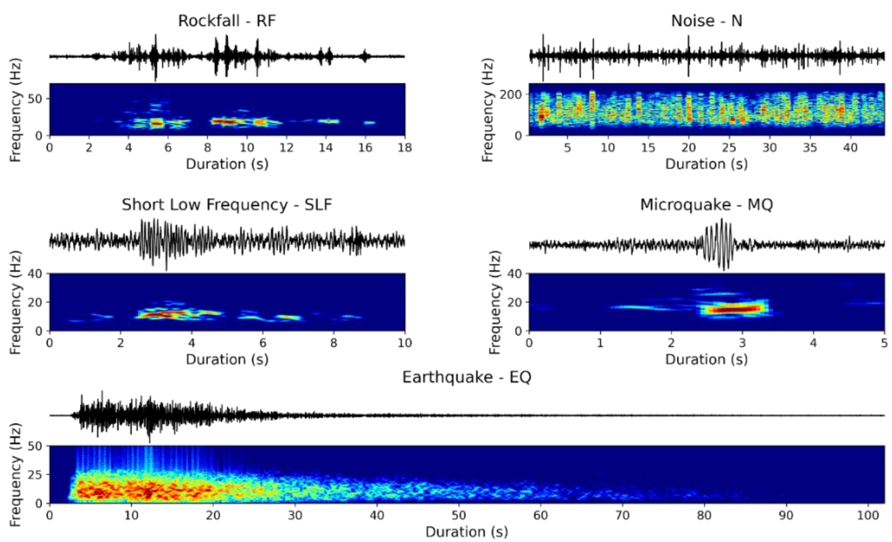
CLASSIFICATION | CONTINUOUS DATA

Dense Nodes Network : Super-Sauze Landslide



- Dense network of 50 seismic stations
- Deployed from the 18th of June, 2016 to the 17th of July, 2016
- **6790** detected **events**
- **5 classes** dominated by noise
- Each event is seen by > 20 stations
- **Strongly unbalanced : > 75% Noise**





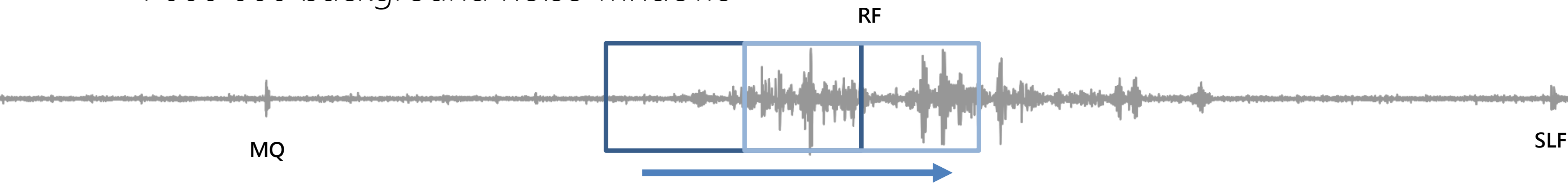
Rimpot et al.

CLASSIFICATION | CONTINUOUS DATA

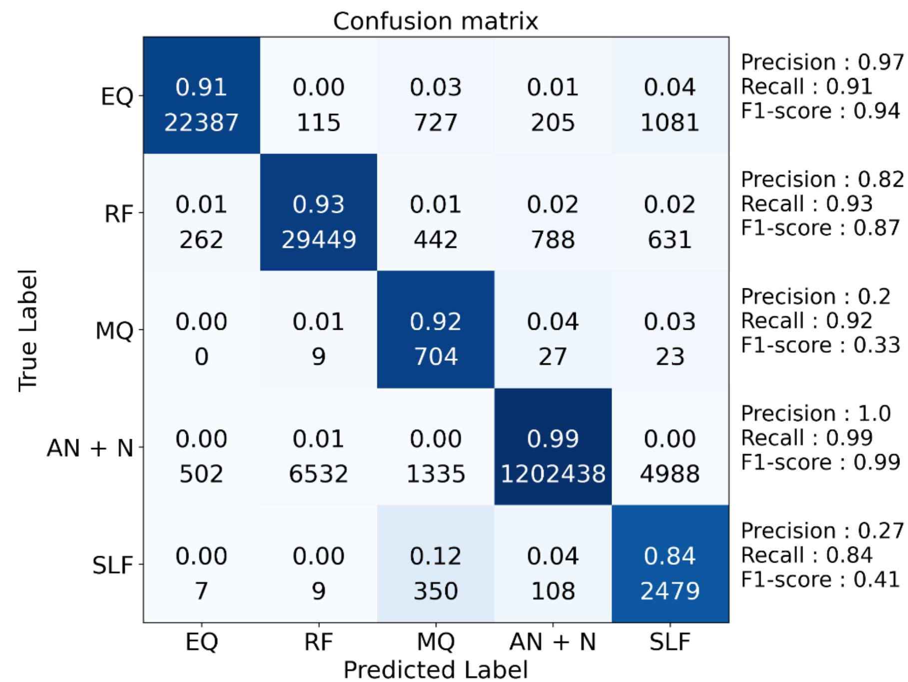
Dataset - Windowed catalogue

Rimpot et al.

- 1s-sliding windows of 18s-length
- + 1 000 000 background noise windows

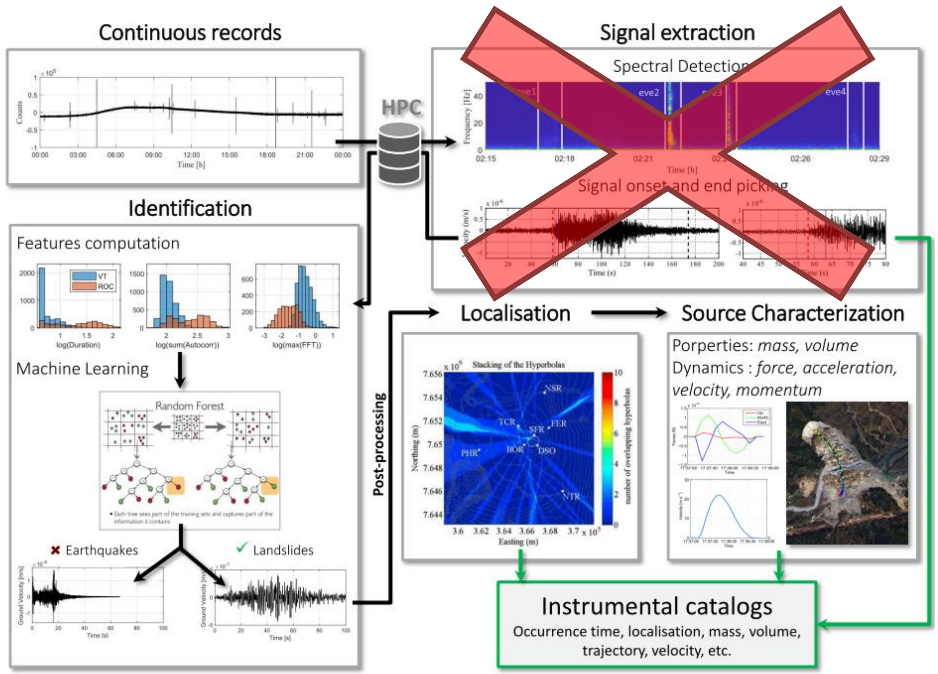


Sub dataset		
Classes	**Nb windows**	**Proportion**
Noise **(N)**	220 798	17,11%
Ambient Noise **(AN)**	1 000 000	77,48%
Total Noise	**1 220 798**	94,59%
Rockfall **(RF)**	34 073	2,64%
Earthquake **(EQ)**	27 017	2,09%
Short Low Frequency **(SLF)**	5 456	0,42%
Microquake **(MQ)**	3 266	0,25%
Total Event	**69 812**	5,41%
Total	**1 290 610**	100,00%



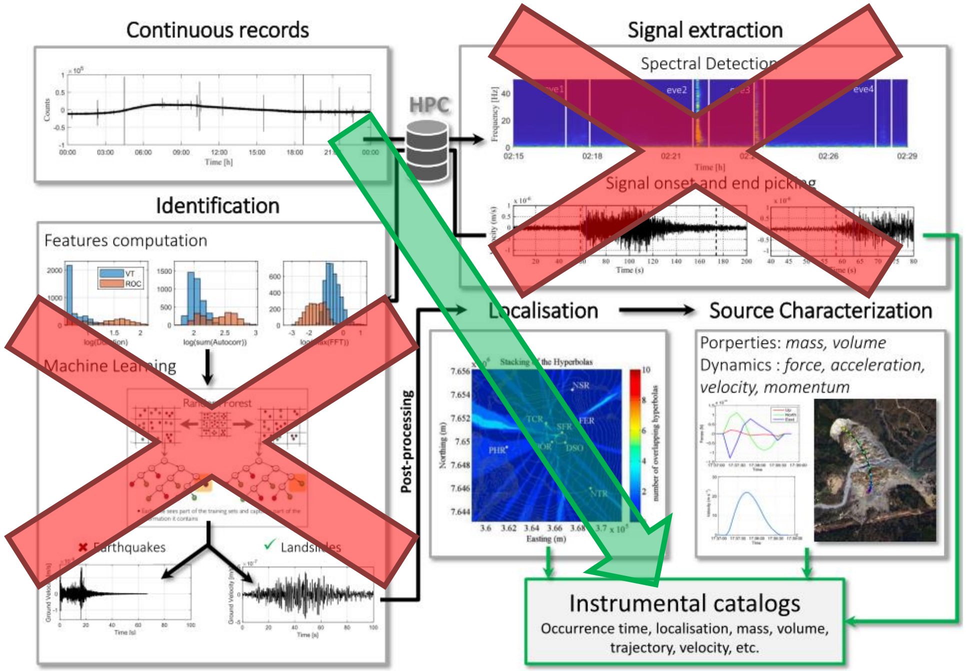
- **XGBoost** on the sub-dataset :
- Trainset : **2500 windows** / Classes

CLASSIFICATION | SELF-SUPERVISED



Can we remove the need to have an *initial catalogue* ?

Manual initial catalogue = subjective, based on a priori knowledge on the classes, not comprehensive = bias



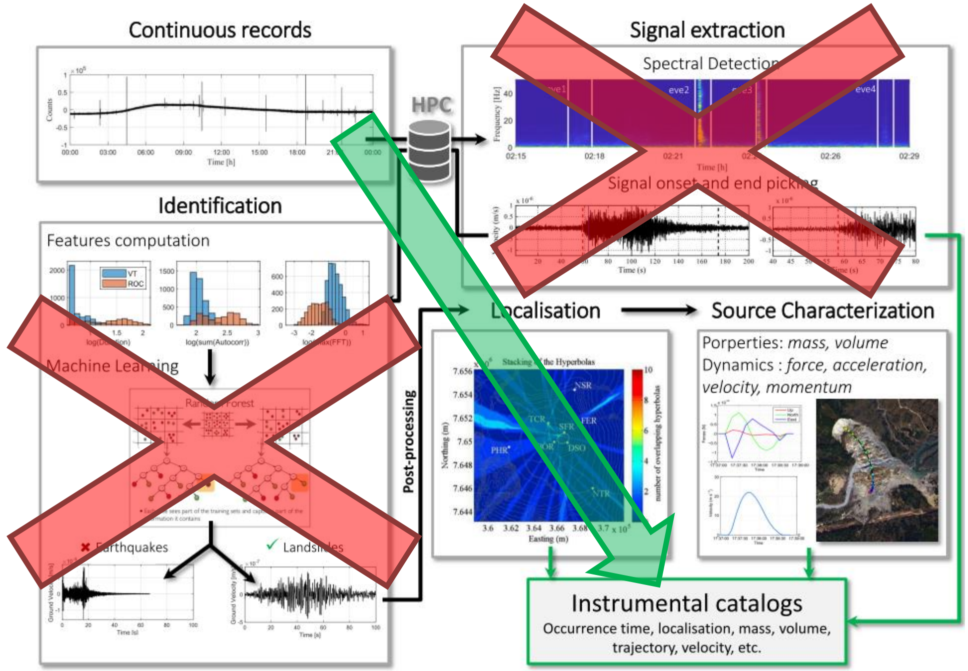
Can we remove the need to have an *initial catalogue* ?

Manual initial catalogue = subjective, based on a priori knowledge on the classes, not comprehensive = bias

Self-supervised learning :

- Needed to processes **unlabelisable datasets**
- Can achieve high scores with **few examples**
- Can find **rare and « exotic » events**

BYOL [Grill et al., 2020], DeepClusterV2, DINO, SwAV [Caron et al., 2020a, 2020b, 2021], MoCo, SimCLR [Chen et al., 2020a, 2020b], ...



CLASSIFICATION | SELF-SUPERVISED

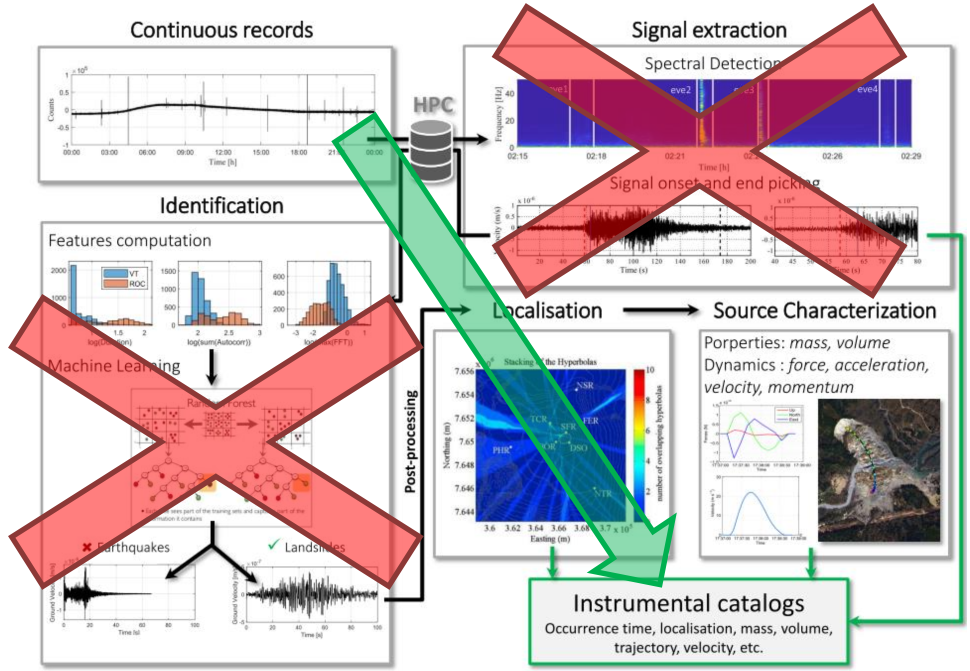
Can we remove the need to have an *initial catalogue* ?

Manual initial catalogue = subjective, based on a priori knowledge on the classes, not comprehensive = bias

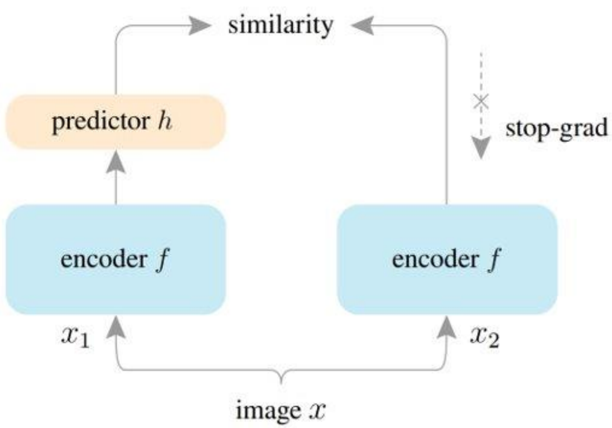
Self-supervised learning :

- Needed to processes **unlabelisable datasets**
- Can achieve high scores with **few examples**
- Can find **rare and « exotic » events**

BYOL [Grill et al., 2020], DeepClusterV2, DINO, SwAV [Caron et al., 2020a, 2020b, 2021], MoCo, SimCLR [Chen et al., 2020a, 2020b], …



Simple Siamese network (SimSiam) [Chen & He, 2021]

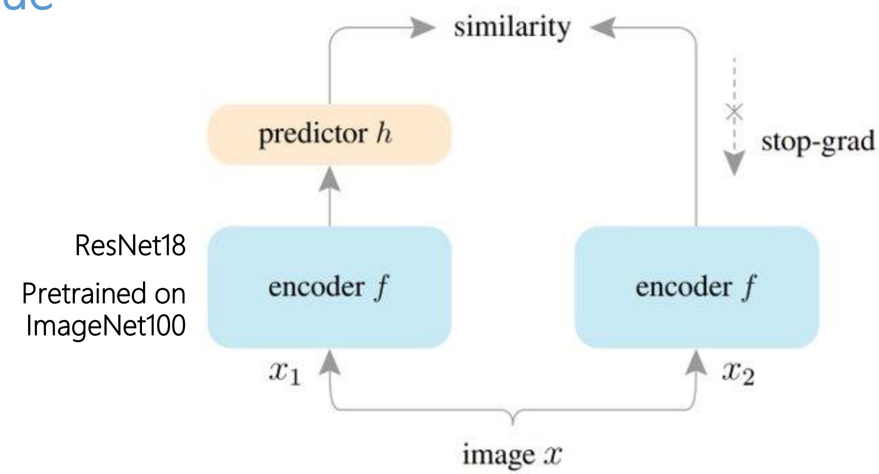


SimSiam +++

- ✓ No need for large batches
- ✓ No need for negative sample pairs

Self-supervised Learning

Simple Siamese network (SimSiam) [Chen & He, 2021]

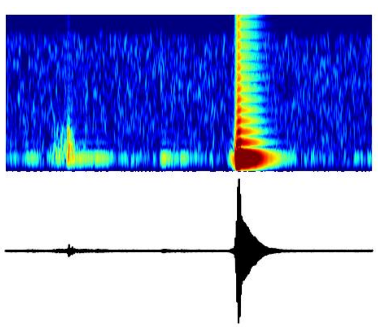


Mayotte volcano - REVOSIMA catalogue

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19

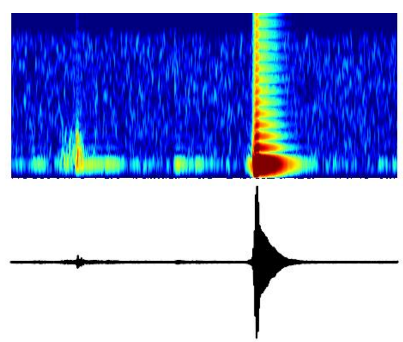
Classes	Nbr Events
Volcano-Tectonic earthquakes (VT)	2 008
Hydro-Acoustic signals (HA)	1 626

HD images

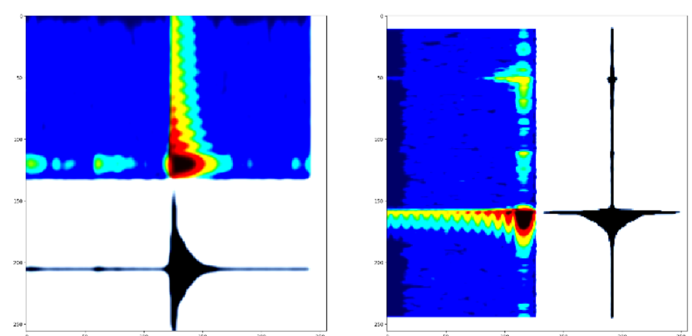


30 s

256 x 256



Data transformation



Self-supervised Learning

Mayotte volcano - REVOSIMA catalogue

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19

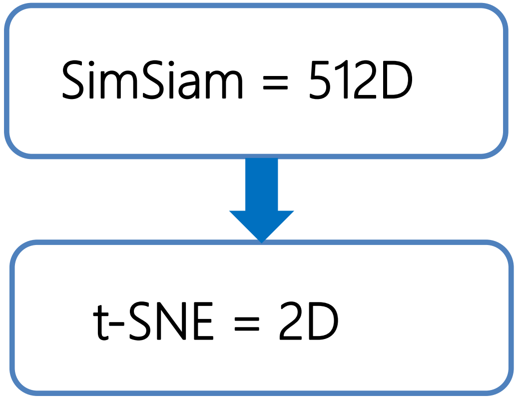
SimSiam = 512D

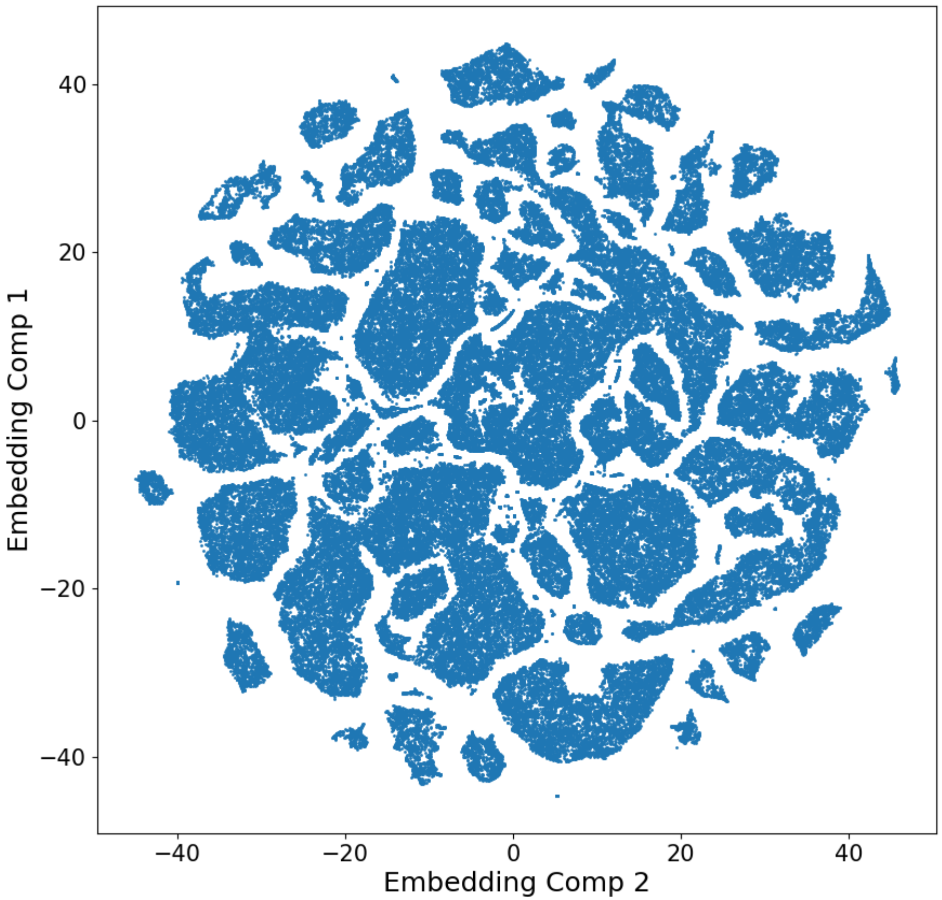
CLASSIFICATION | SELF-SUPERVISED

Self-supervised Learning

Mayotte volcano - REVOSIMA catalogue

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19





t-distributed stochastic neighbor embedding [Van der Maaten & Hinton, 2008]

CLASSIFICATION | SELF-SUPERVISED

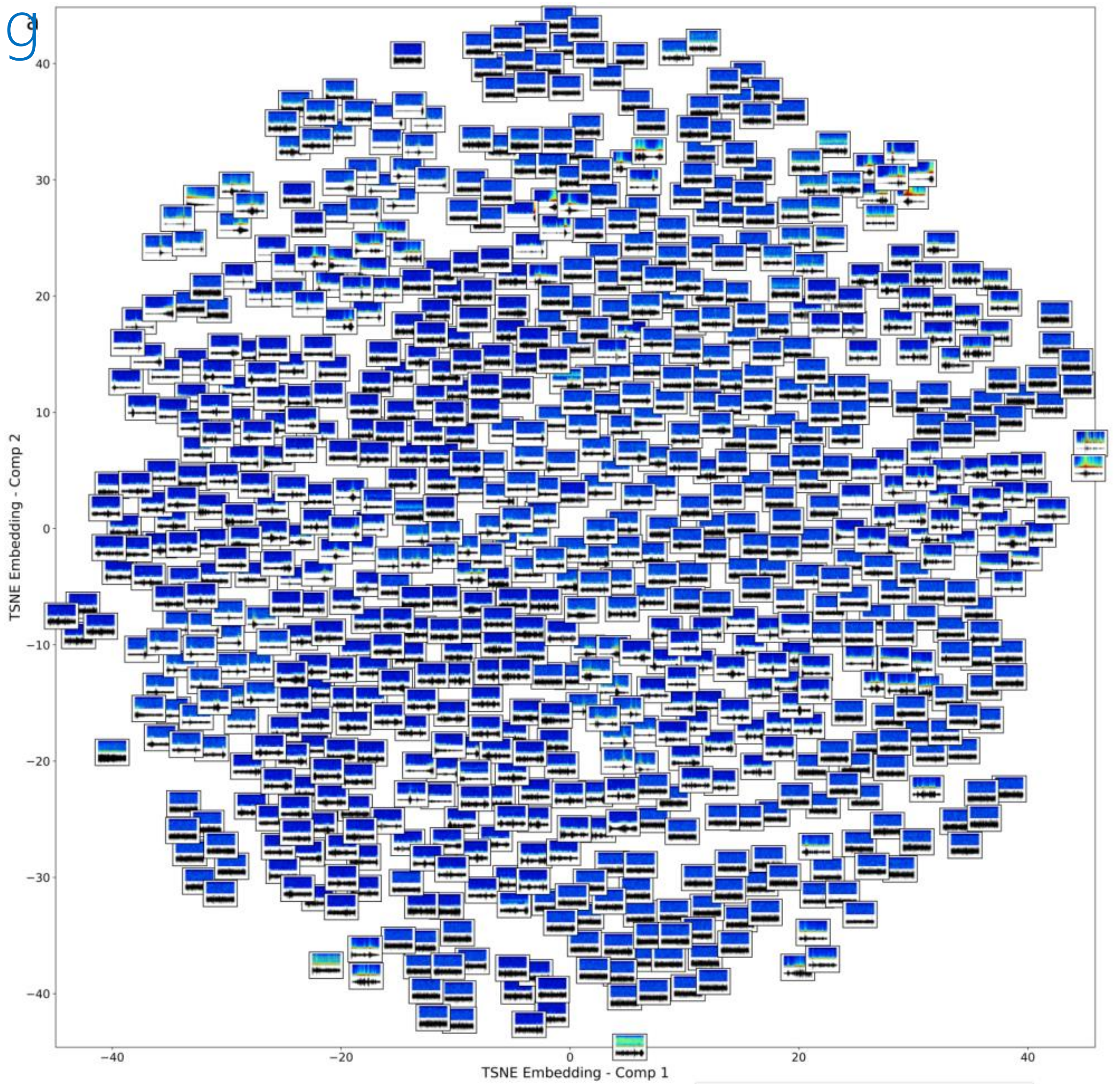
Self-supervised Learning

Mayotte volcano - REVOSIMA

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19

SimSiam = 512D

t-SNE = 2D



t-distributed stochastic neighbor embedding [Van der Maaten & Hinton, 2008]

CLASSIFICATION | SELF-SUPERVISED

Self-supervised Learning

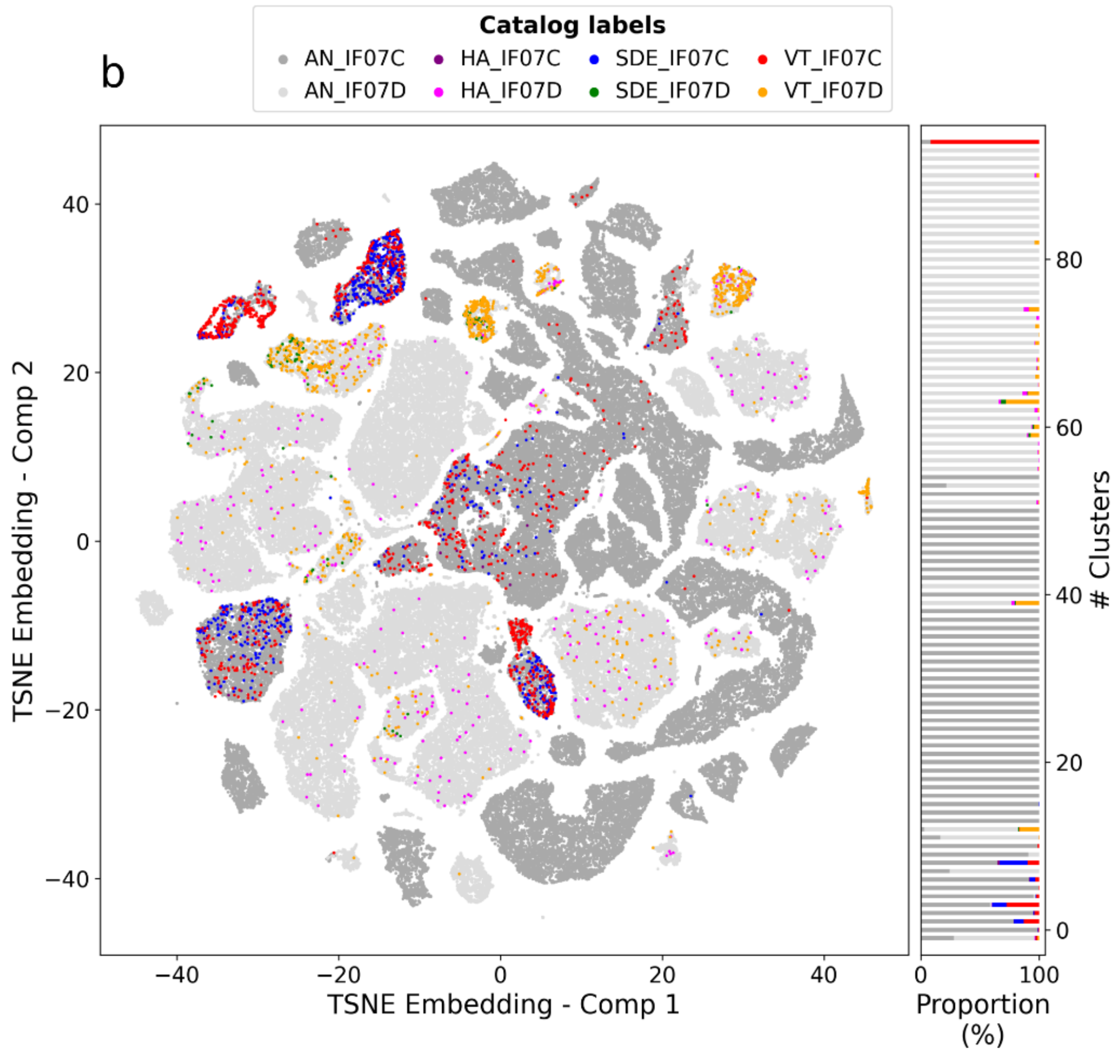
Mayotte volcano - REVOSIMA cat

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19

SimSiam = 512D

↓

t-SNE = 2D



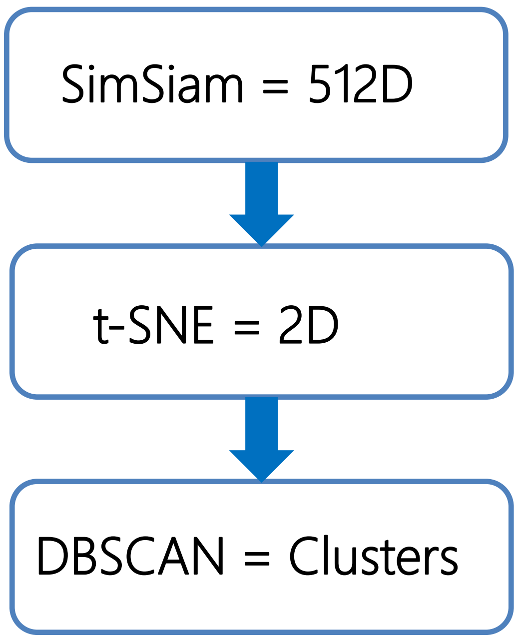
t-distributed stochastic neighbor embedding [Van der Maaten & Hinton, 2008]
density-based spatial clustering of applications with noise [Ester *et al.*, 1996]

CLASSIFICATION | SELF-SUPERVISED

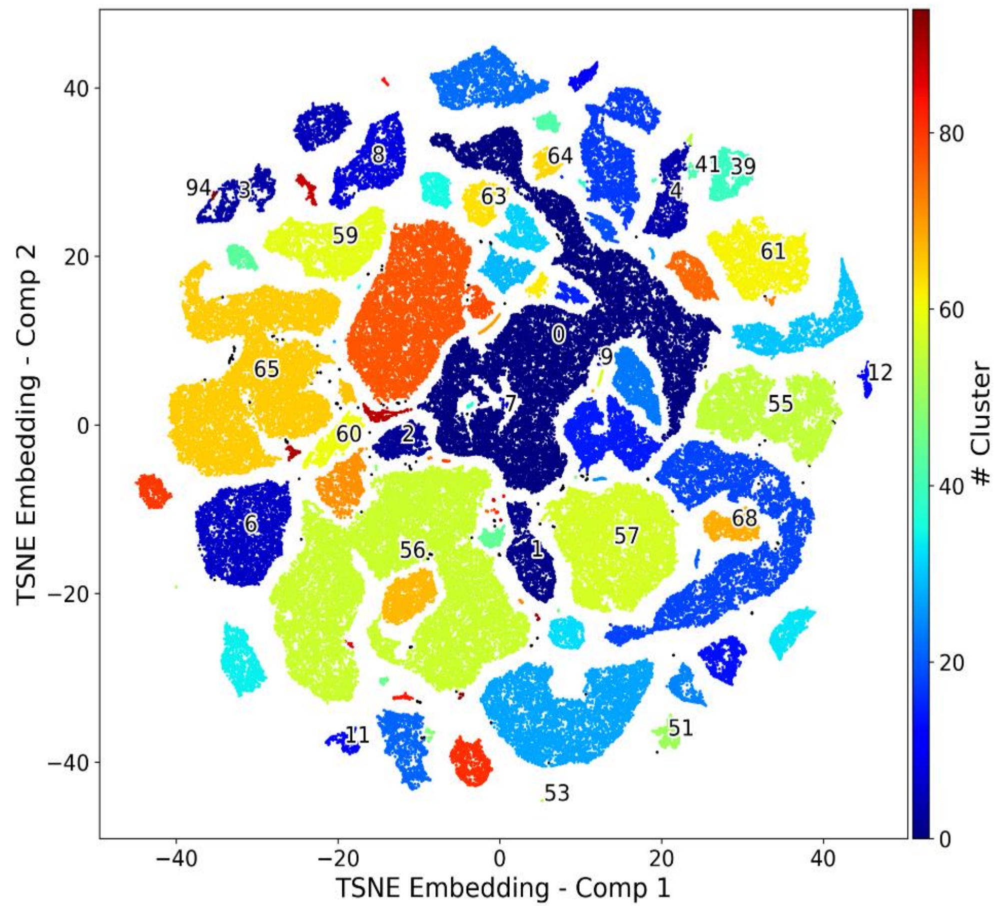
Self-supervised Learning

Mayotte volcano - REVOSIMA cat

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19



a



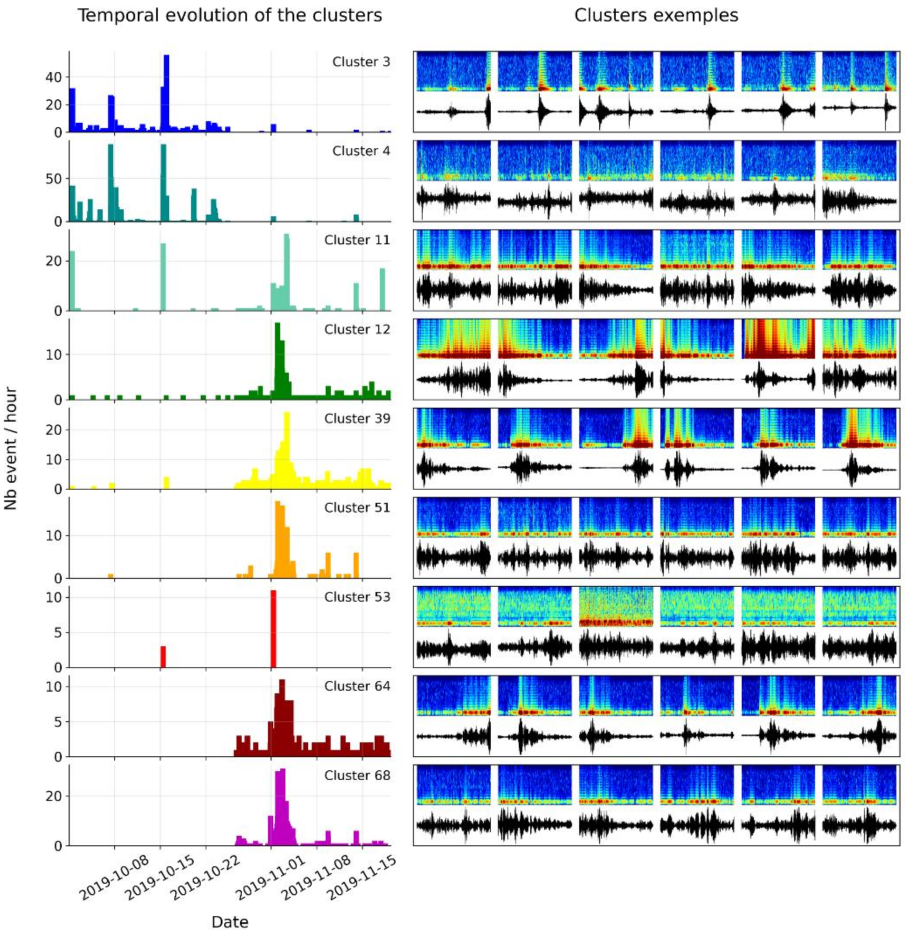
t-distributed stochastic neighbor embedding [Van der Maaten & Hinton, 2008]
density-based spatial clustering of applications with noise [Ester *et al.*, 1996]

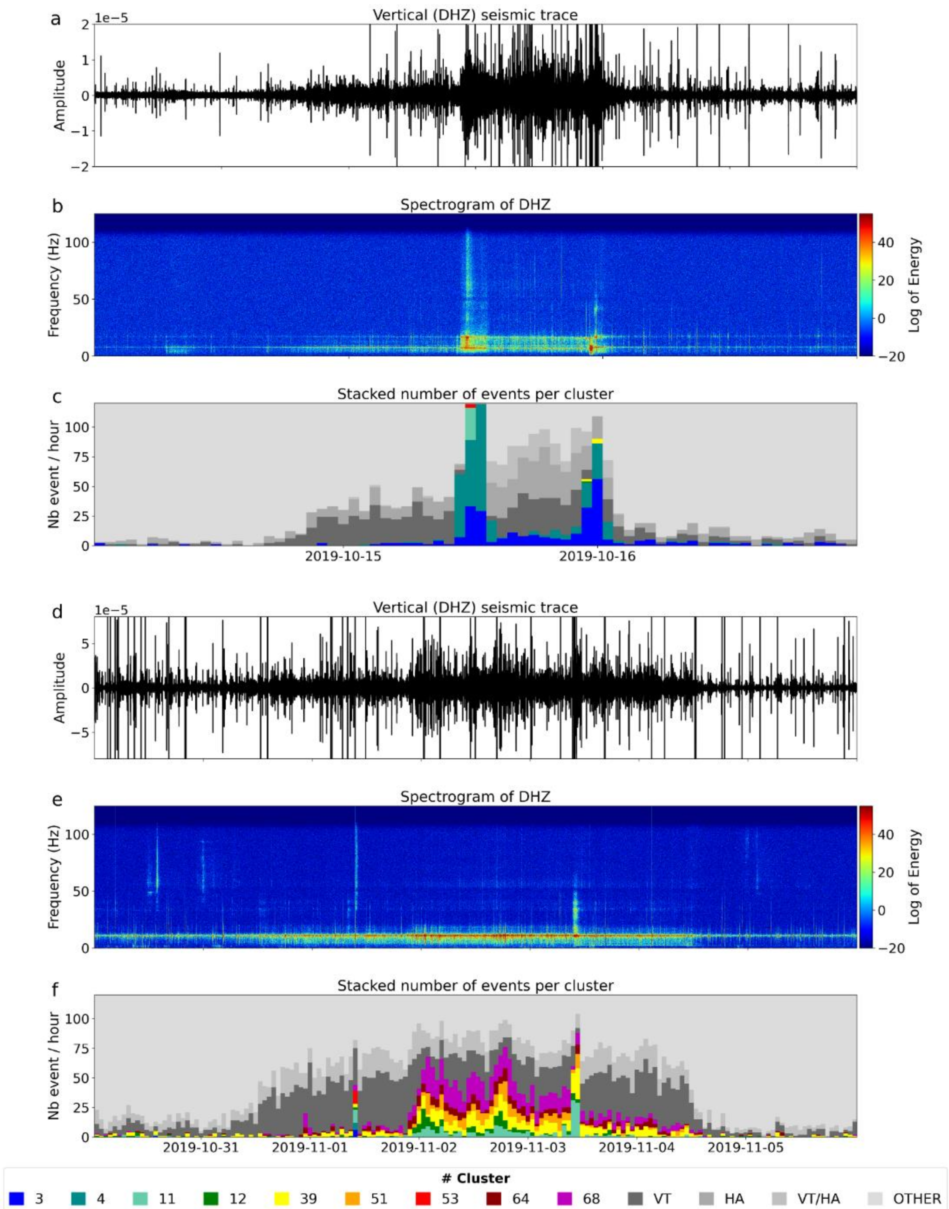
CLASSIFICATION | SELF-SUPERVISED

Self-supervised Learning

Mayotte volcano - REVOSIMA catalogue

- 2 stations : IF07C & IF07D
- From 1/10/19 to 19/11/19

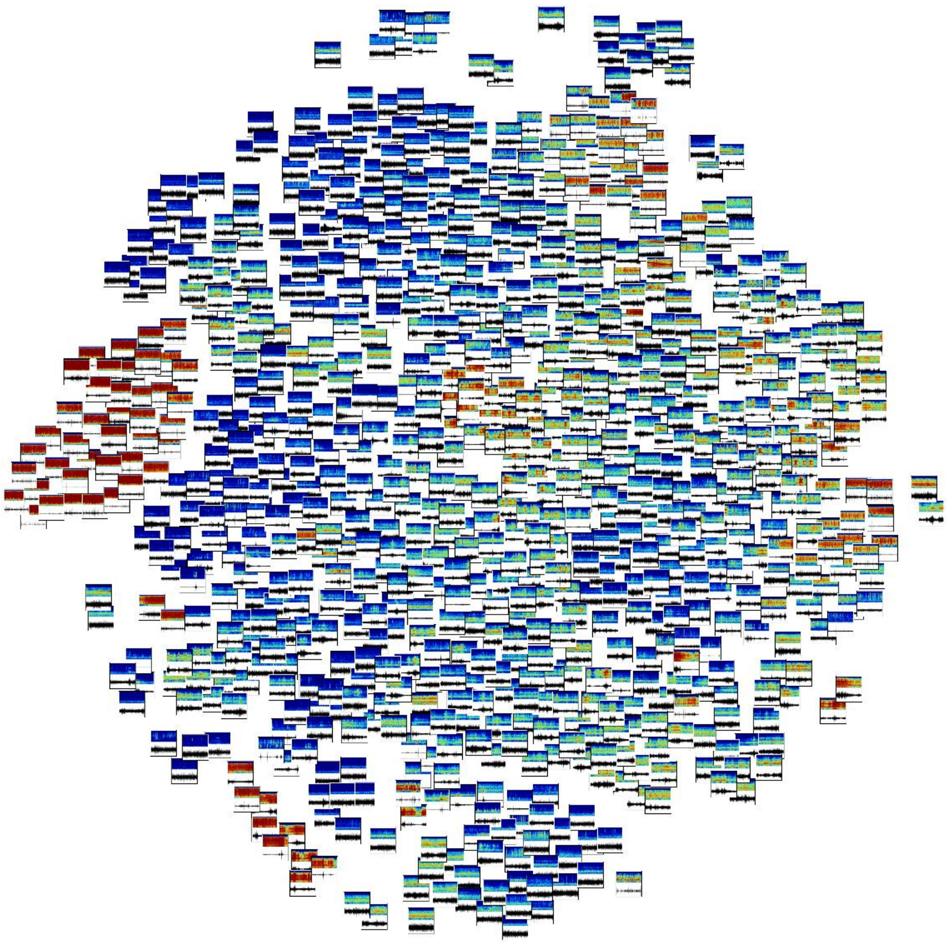




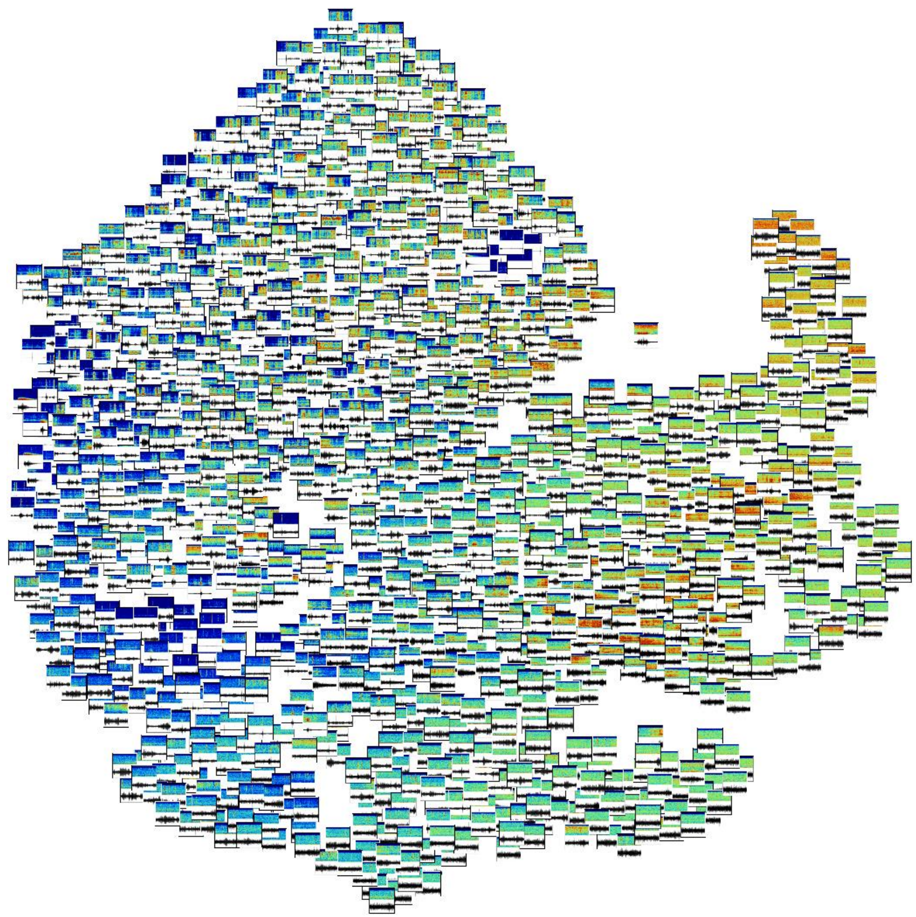
CLASSIFICATION | SELF-SUPERVISED

Self-supervised Learning

Marie sur Tinée – landslide



Plan de Chauvet – rock glacier



CONCLUSIONS :

- ✓ SSL able to process continuous seismic data
- ✓ SSL able to reconstruct and improve existing catalogs
- ✓ SSL able to find rare events

SSL = synoptic and comprehensive view of a dataset

WIP :

- Multistations
- Remove the need to transform the data to images

CHALLENGES :

- A global pretrained model for seismological data ?
- How to apply this to large volume (years, nodes, DAS) ? > VRE

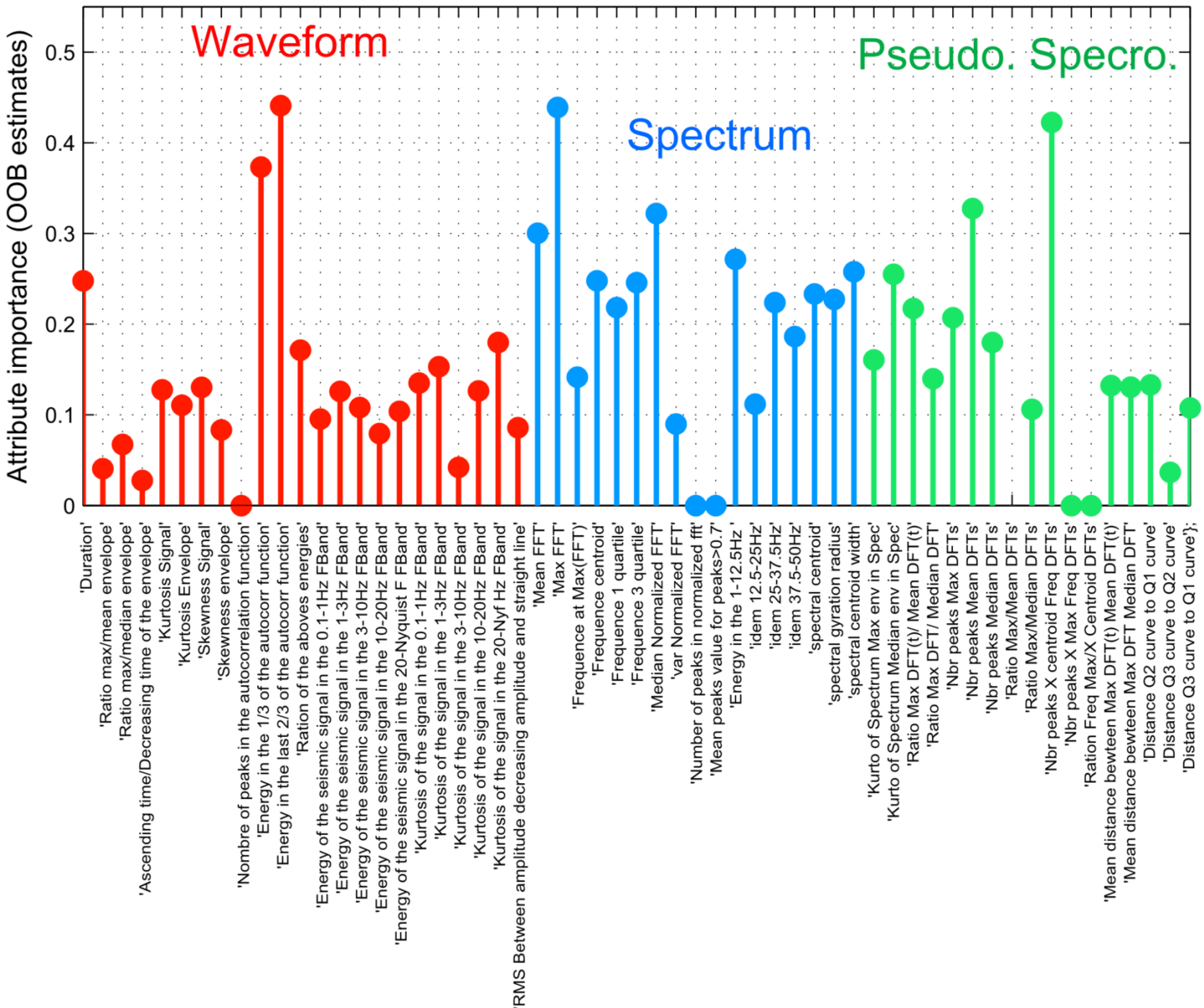
EaSy Data
Earth System Data Repository

DATATERRA

EPOS-FR
EUROPEANPLATEOBSERVINGSYSTEM – FRANCE

Thank you!

Contact : hibert@unistra.fr



Waveform
Spectrum
Pseudo. Specro.
Attribute importance (OOB estimates)
0
0.1
0.2
0.3
0.4
0.5
'Duration'
'Ratio max/mean envelope'
'Ratio max/median envelope'
'Ascending time/Decreasing time of the envelope'
'Kurtosis Signal'
'Kurtosis Envelope'
'Skewness Signal'
'Skewness envelope'
'Nombre of peaks in the autocorrelation function'
'Energy in the 1/3 of the autocorr function'
'Energy in the last 2/3 of the autocorr function'
'Ration of the aboves energies'
'Energy of the seismic signal in the 0.1-1Hz FBand'
'Energy of the seismic signal in the 1-3Hz FBand'
'Energy of the seismic signal in the 3-10Hz FBand'
'Energy of the seismic signal in the 10-20Hz FBand'
'Energy of the seismic signal in the 20-Nyquist F FBand'
'Kurtosis of the signal in the 0.1-1Hz FBand'
'Kurtosis of the signal in the 1-3Hz FBand'
'Kurtosis of the signal in the 3-10Hz FBand'
'Kurtosis of the signal in the 10-20Hz FBand'
'Kurtosis of the signal in the 20-Nyf Hz FBand'
'RMS Between amplitude decreasing and straight line
'Mean FFT'
'Max FFT'
'Frequence at Max(FFT)'
'Frequence centroid'
'Frequence 1 quartile'
'Frequence 3 quartile'
'Median Normalized FFT'
'var Normalized FFT'
'Number of peaks in normalized fft'
'Mean peaks value for peaks>0.7'
'Energy in the 1-12.5Hz '
'idem 12.5-25Hz'
'idem 25-37.5Hz'
'idem 37.5-50Hz'
'spectral centroid'
'spectral gyration radius'
'spectral centroid width'
'Kurto of Spectrum Max env in Spec'
'Kurto of Spectrum Median env in Spec'
'Ratio Max DFT(t)/ Mean DFT(t)'
'Ratio Max DFT/ Median DFT'
'Nbr peaks Max DFTs'
'Nbr peaks Mean DFTs'
'Nbr peaks Median DFTs'
'Ratio Max/Mean DFTs'
'Ratio Max/Median DFTs'
'Nbr peaks X centroid Freq DFTs'
'Nbr peaks X Max Freq DFTs'
'Ration Freq Max/X Centroid DFTs'
'Mean distance bewteen Max DFT(t) Mean DFT(t)'
'Mean distance bewteen Max DFT Median DFT'
'Distance Q2 curve to Q1 curve'
'Distance Q3 curve to Q2 curve'
'Distance Q3 curve to Q1 curve'};

Testing the RF algorithm + Feature in different contexts

Local scale :

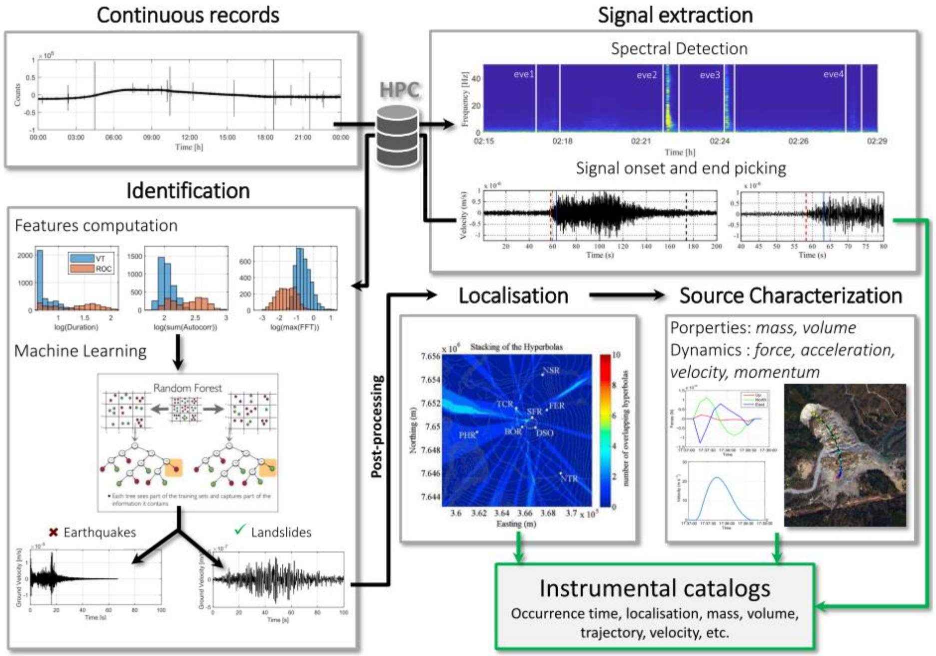
- Super-Sauze [*Provost et al., 2017*] – 4 classes, ~900 eve. Success rate : 90%
- Piton de la Fournaise volc. [*Maggi et al., 2017; Hibert et al., 2017*] – 2-8 classes, 13000+ eve. : 90-95+%
- La Clapière – 4 classes, ~11100 eve. : 92%
- Séchilienne – 4 classes, ~130000 eve. : 91%
- Knipovich Ridge [*Domel et al., in press*] : 87%

Regional scale :

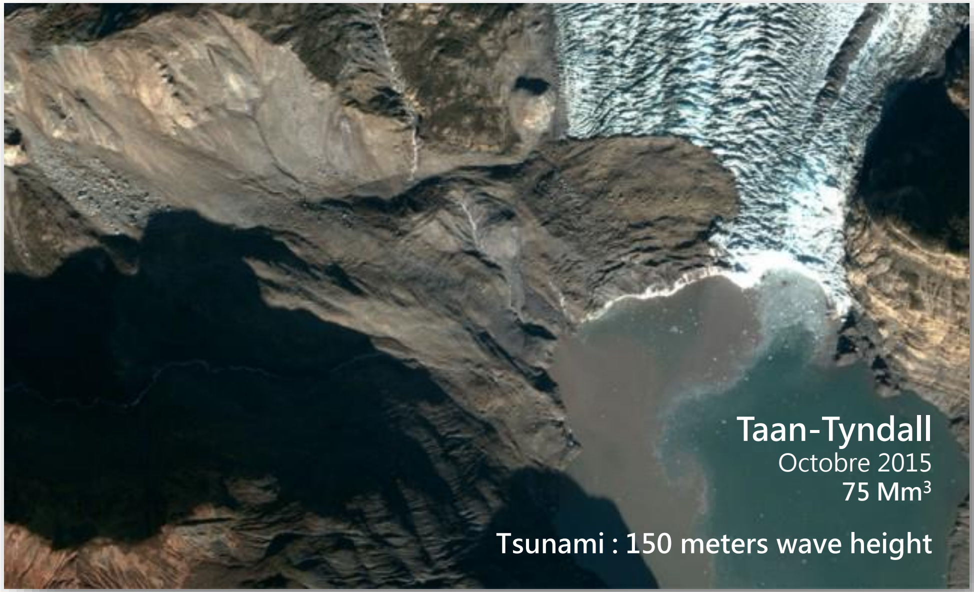
- Alaska [Hibert et al., 2019]
- Alps : WIP [Groult et al., in prep.] > ANR HighLand
- Greenland [Pirot et al., in prep.]

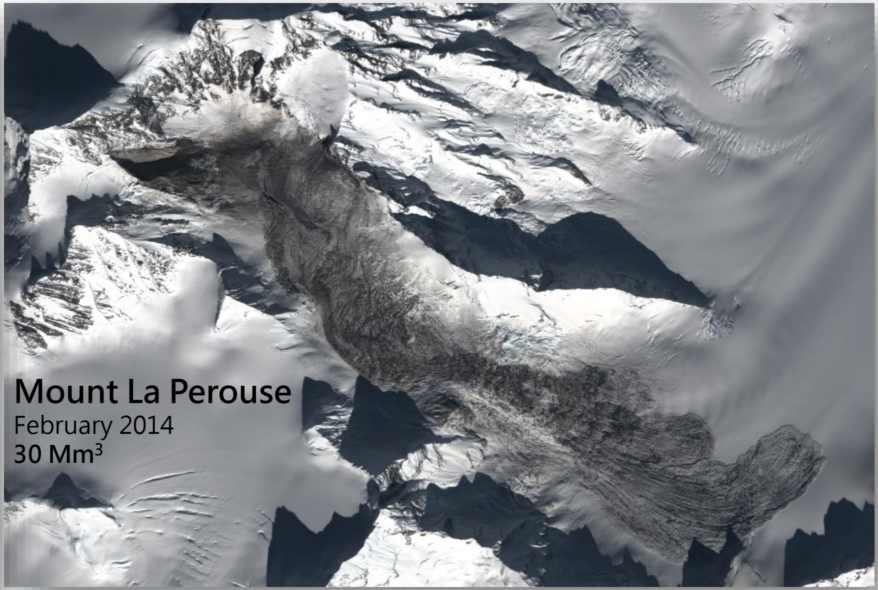
Processing streams of data :

- Illgraben/Piz Cengalo [*Wenner et al., 2021; Chmiel et al., 2021*]: 80-90%
- DAS [*Huynh et al., 2022*] : 87%
- Super-Sauze [*Rimpot et al, in prep.*]



CLASSIFICATION | LANDSLIDES





Scientific question : How is climate change impacting landslides activity in high latitude/altitude regions of the world ?

> Need for comprehensive catalogues of landslides

CLASSIFICATION | ALASKA

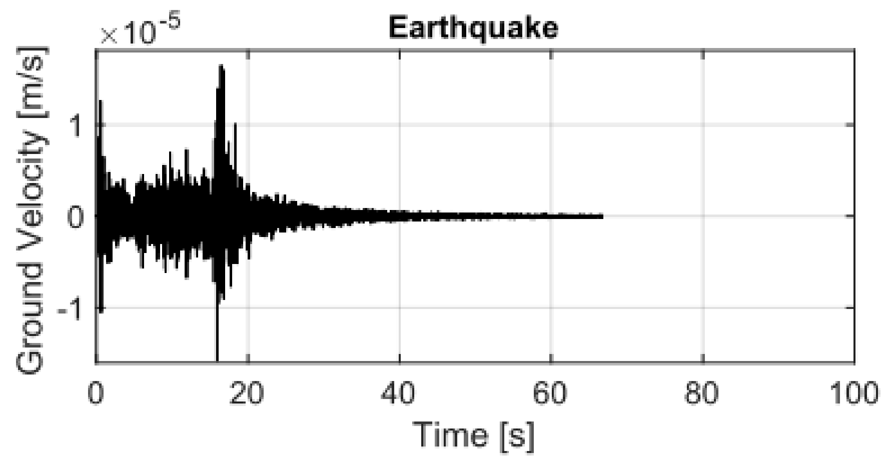
Training Set : 2 classes

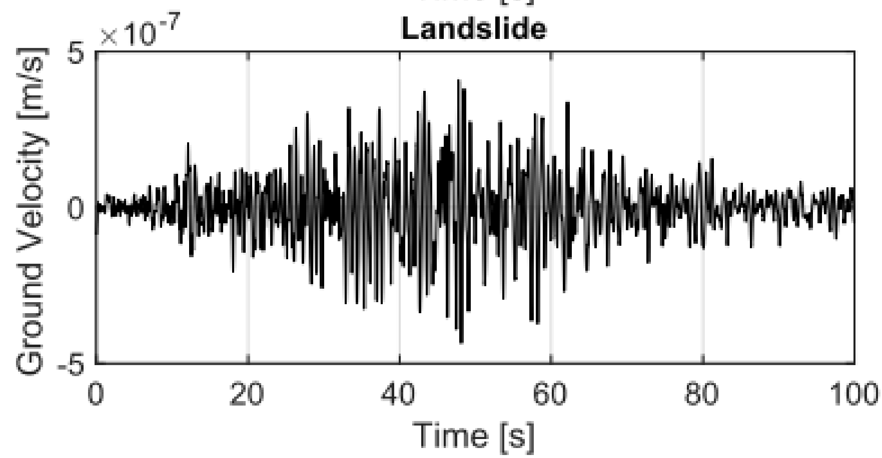
Earthquakes :

- **290** Earthquakes recorded by the Alaskian network (AK) in january 2016 (M 2.5-7.1)
- **3636** HF seismic signals recorded by 124 stations

Landslides :

- **11** landslides (Volume>1Mm^3)
- **205** HF seismic signals recorded
- Events known or seismically detected (GCMT project, *Ekström et al.*)





Algorithm implementation

Tests performed : 100 iterations of training the algorithm with a sub-set of the training set and then identification of the rest of the set

Signal Approach :

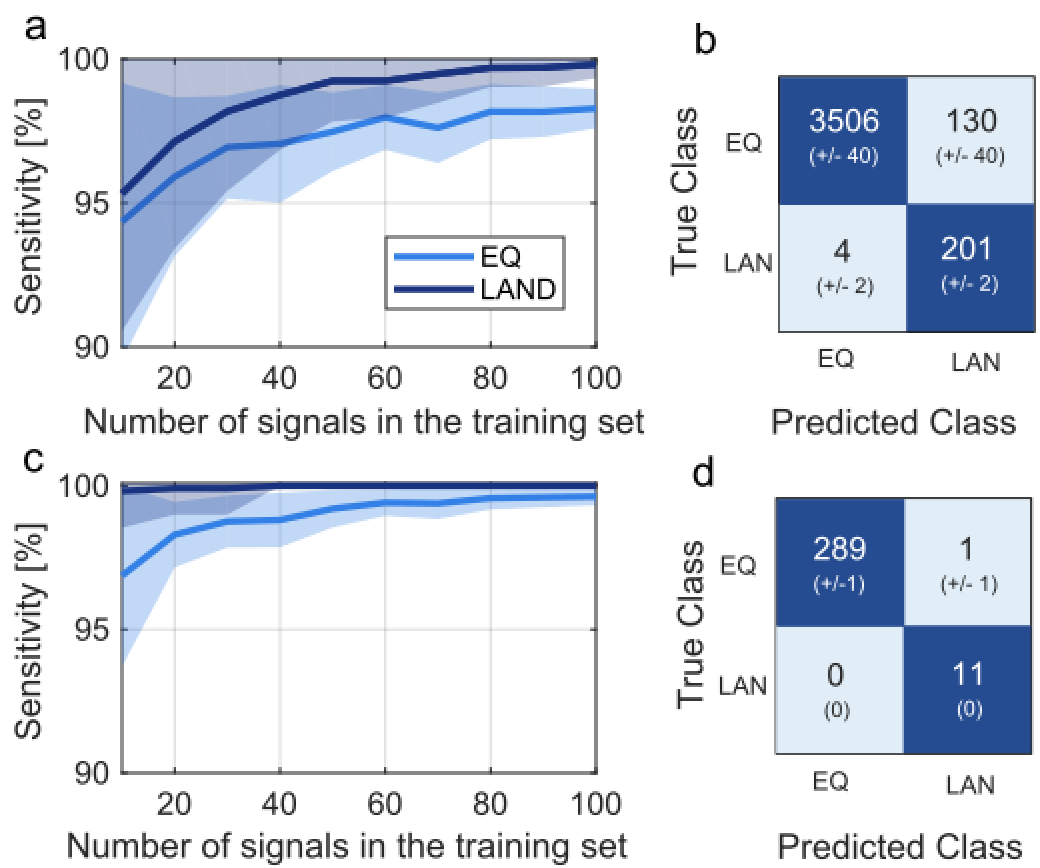
Identifying one event from one signal

Accuracy : 98%
But high rate of **false alarm**!

Event Approach :

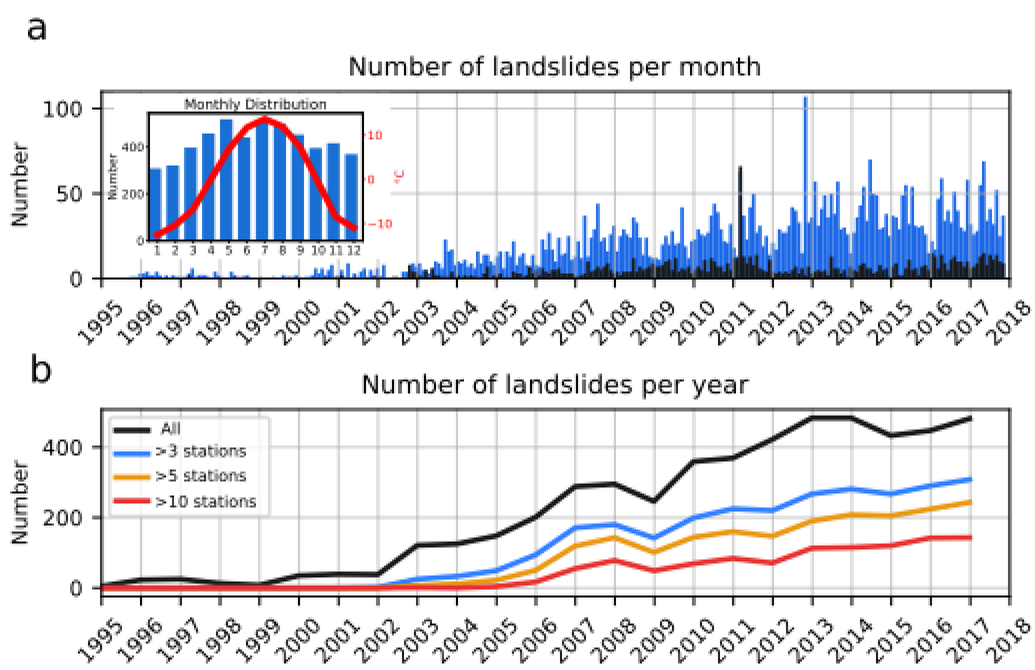
Identifying one event from the vote casted by each signal (+score) associated with the event

Accuracy : 99%
Worst case : 1 EQ identified as landslide.
No landslides missed



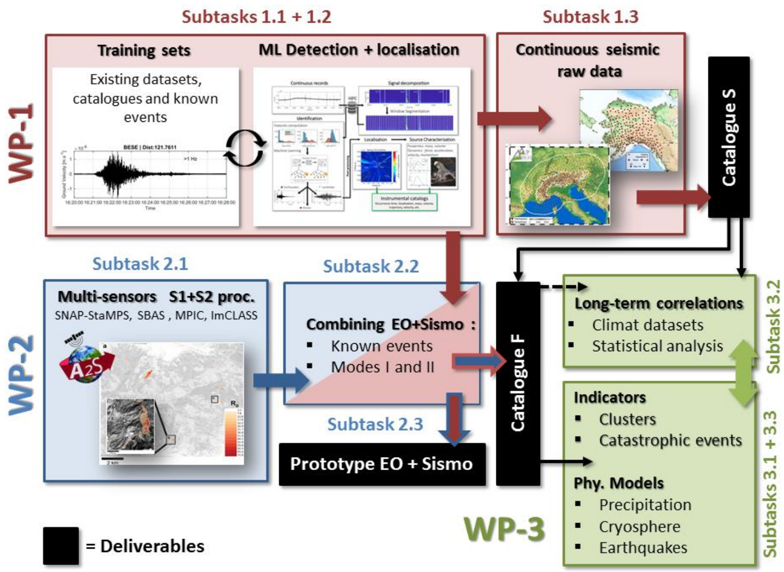
Application to 22 years of continuous data

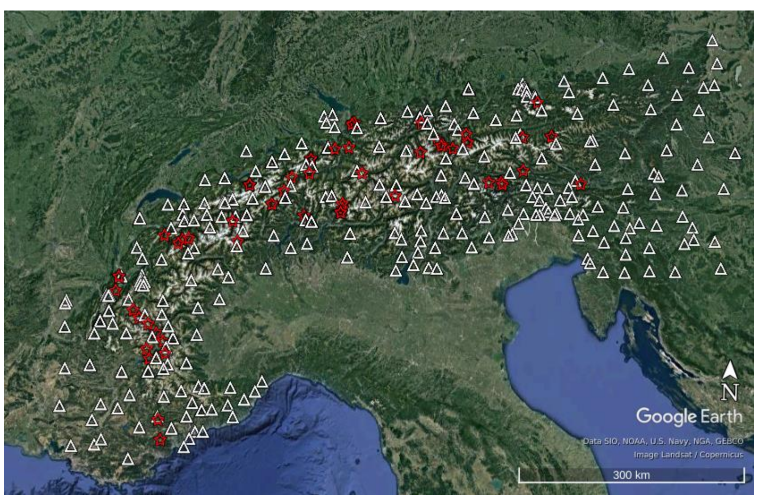
- HPC implementation : **10h of processing for 240+ stations** (~12 months on a laptop)
- Zone of detection: **20° x 20° -** Lat: 48°/68°, Lon: -124°/-144°
- **6213** potential landslide detections on **more than 1 station**, 5087 (82%) landslides confirmed by manual inspection of the signals
- **All** of previously known landslides have been detected



CLASSIFICATION – WIP | ALPS

anr





ANR HighLand

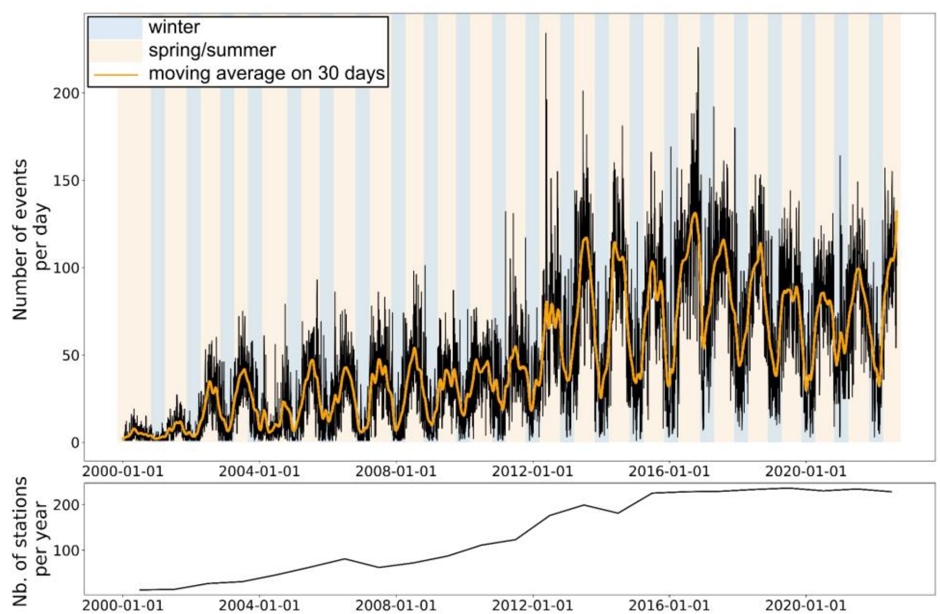
Multi-disciplinary :

- Seismology
- Remote-Sensing
- I.A.

Instrumental Catalogues :

- Date, localization, mass and volume
- In short/near real time
- Retrospectively over 20 years

Groult et al.



Testing the RF algorithm + Feature in different contexts

Local scale :

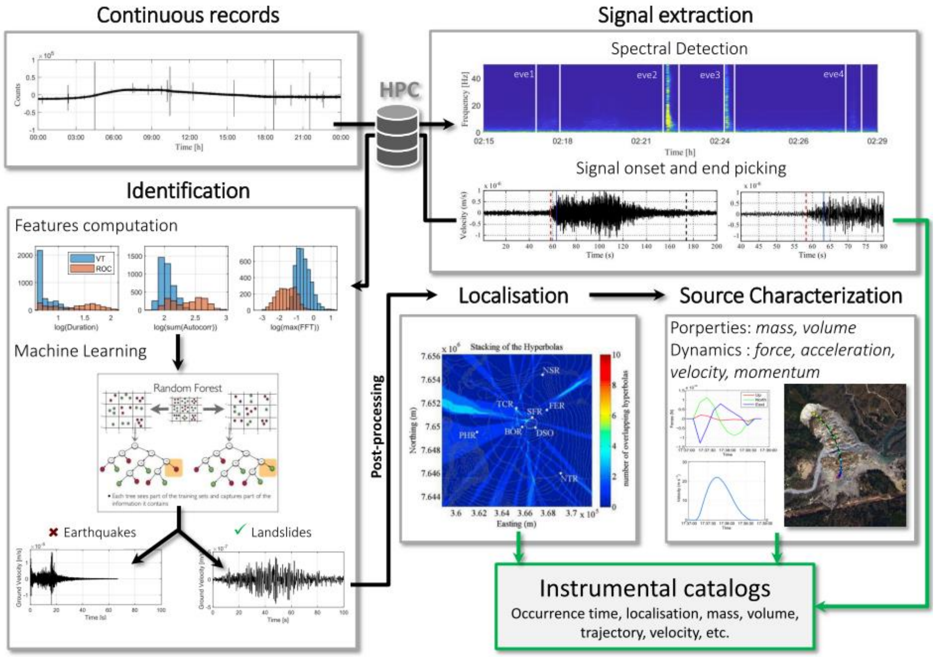
- Super-Sauze [*Provost et al., 2017*] – 4 classes, ~900 eve. Success rate : 90%
- Piton de la Fournaise volc. [*Maggi et al., 2017; Hibert et al., 2017*] – 2-8 classes, 13000+ eve. : 90-95+%
- La Clapière – 4 classes, ~11100 eve. : 92%
- Séchilienne – 4 classes, ~130000 eve. : 91%
- Knipovich Ridge [*Domel et al., in press*] : 87%

Regional scale :

- Alaska [*Hibert et al.*, 2019]
- Alps : WIP [*Groult et al., in prep.*] > ANR HighLand
- Greenland [*Pirot et al., sub.*]

Processing streams of data :

- Illgraben/Piz Cengalo [*Wenner et al., 2021; Chmiel et al., 2021*]: 80-90%
- DAS [*Huynh et al., 2022*] : 87%
- Super-Sauze [*Rimpot et al, in prep.*]



CLASSIFICATION | **GREENLAND**

Why study glacier calving in Greenland ?

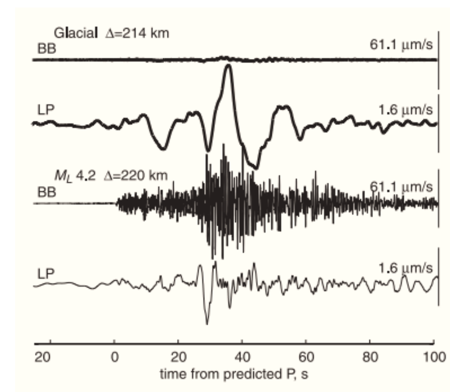
- Indicators of rapid change in the Arctic
- Strong impact on the dynamics/kinematics of these glaciers
- What contribution to ice mass loss and sea level rise?

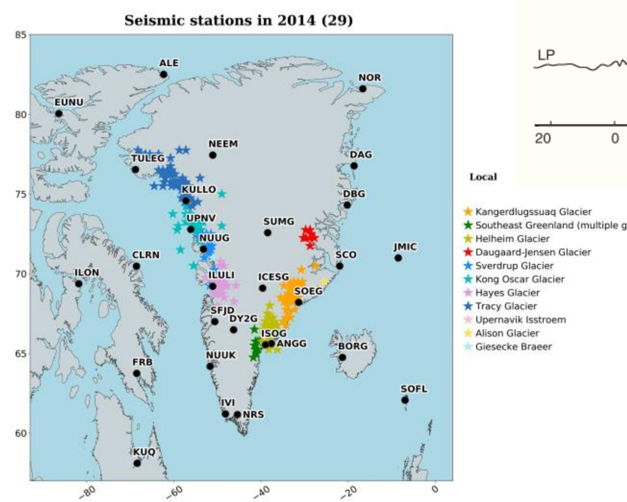
GCMT [Ekström et al.] : first catalogue
1993 – 2013 : 444 Glacial Earthquakes
Ms > 4.5

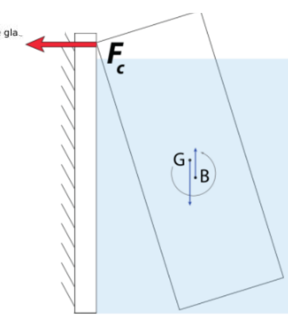
Events Ms < 4.5 not detected

Need for a comprehensive catalogue to address the quantification of ice sheet mass loss

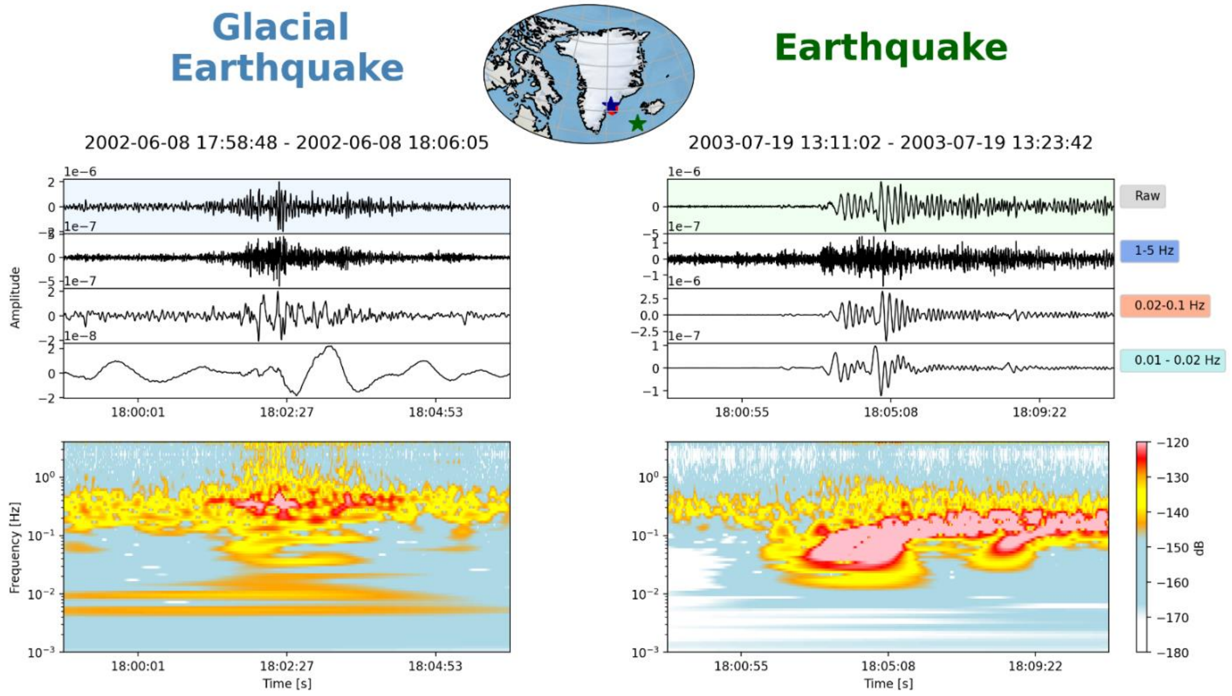
Ekström, Nettles and Abers (2003), Tsai and Ekström (2007), Nettles and Ekström (2010), Sergeant et al. (2016)







Training set : 2 classes



Pirot et al.

GEQ :

- **444** GEQ (M > 4.5)
- **3424** signals
- Known events (GCMT project, *Ekström et al.*)

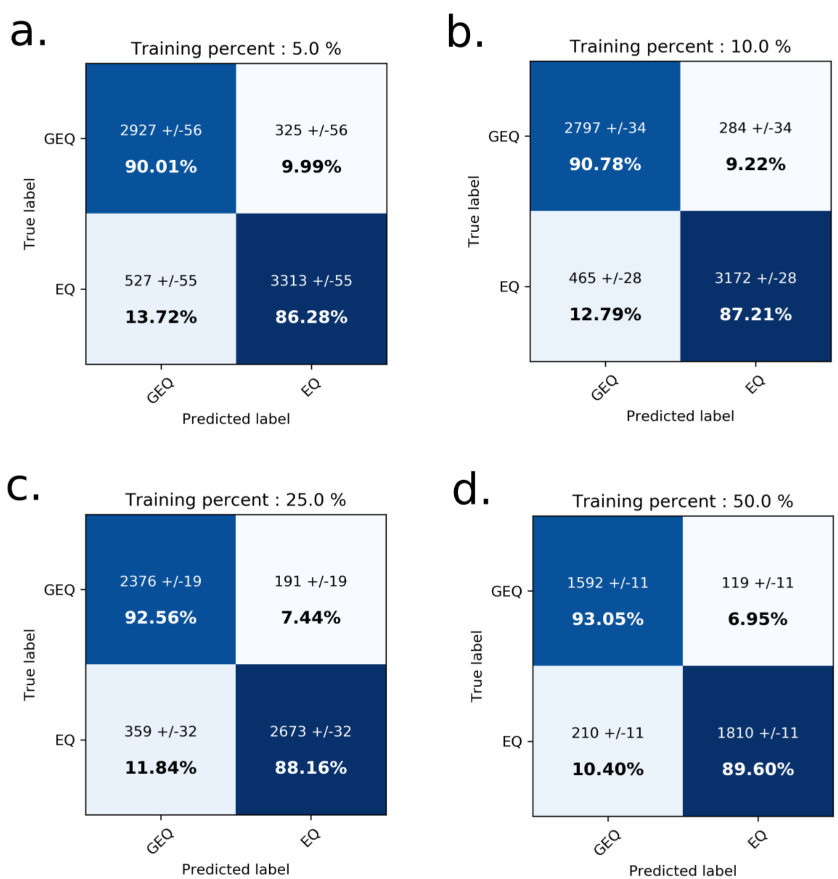
Earthquakes :

- **400** earthquakes recorded by the GLISN network : 1993 to 2013 (Mw 2.5-7.1)
- **4042** signals

CLASSIFICATION | GREENLAND

Application to the GLISN network on 844 days

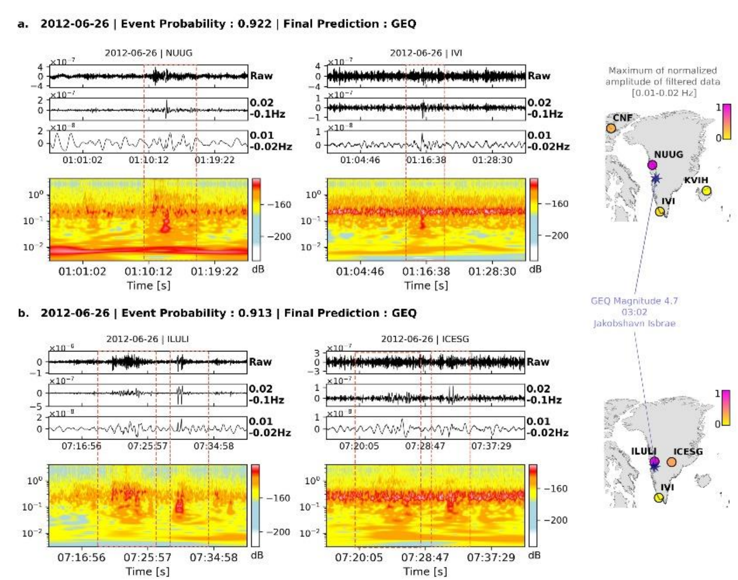
Pirot et al.



- 5791 events > **1670 new GEQ confirmed manualy**

  = 4x the GCMT Cat.

- Events discarded : 758 EQ, possible + GEQ but with signal only on one station



How can seismology help to understand environmental processes ?

- **Detection** and **identification** of active areas (*where? what?*)
- **Monitoring** to alert on possible risks (*when?*)
- **Understanding** the influence of different forcings (meteorological, climatic, tectonic) (*why?*)

Detection & localisation of seismic sources :

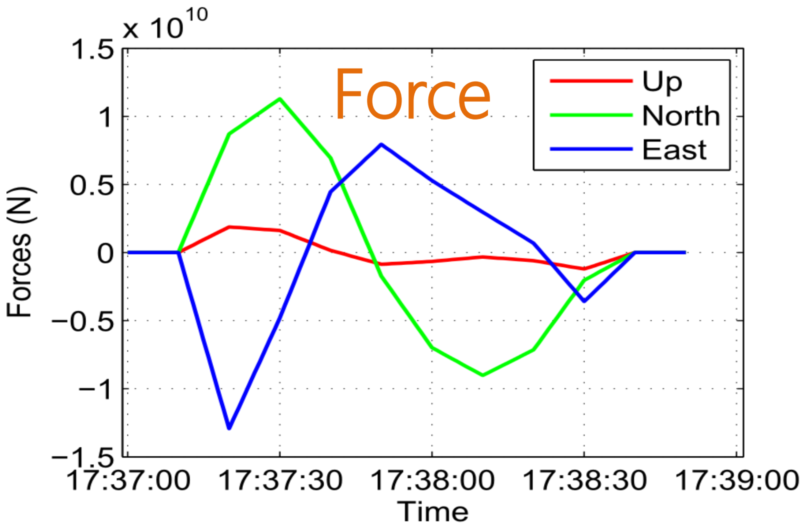
- **Global Scale** : large events (landslides, calving events, etc.)
- **Regional and local scale** : rockfalls, lahars, debris flows, avalanches
- **Endogeneous seismicity** : landslides, glaciers, etc.

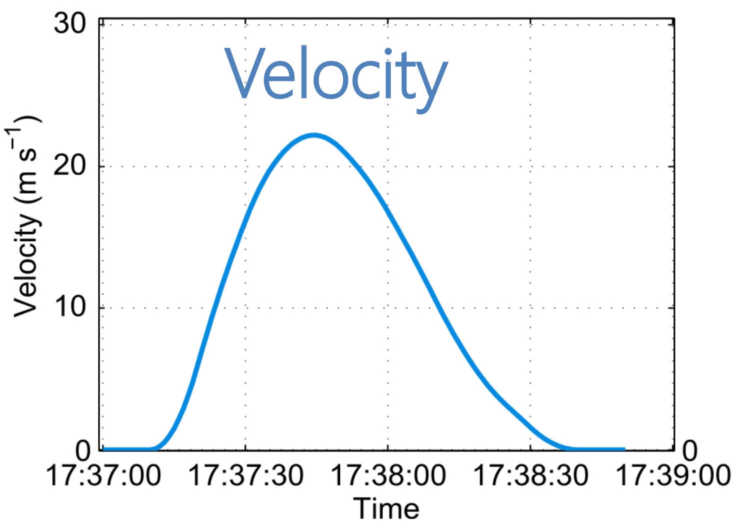
Characterization of the properties and dynamics of the sources :

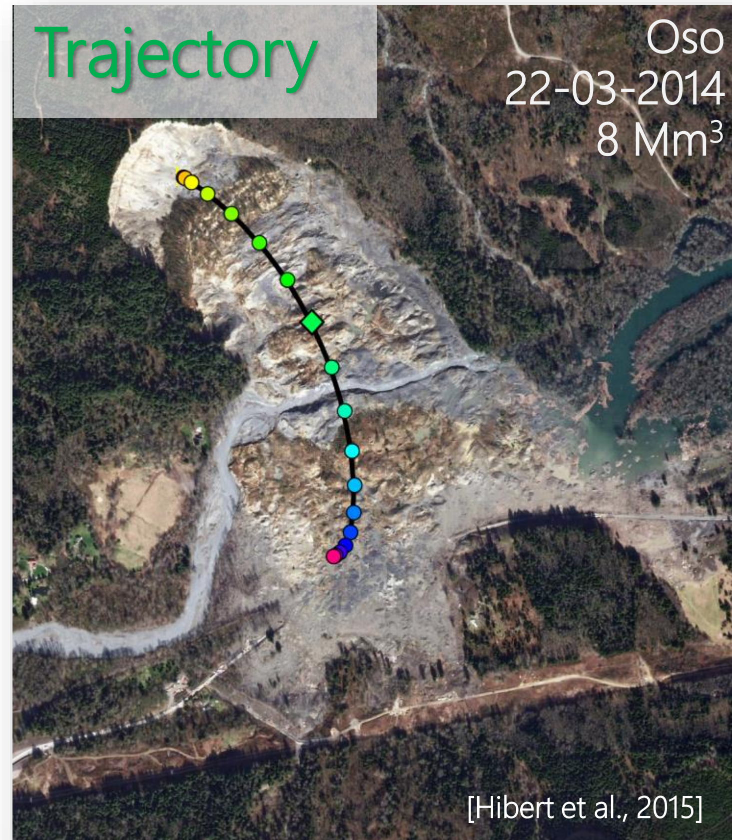
- Inversion and modelisation with **long period waves (>30-40 s)**
- Statistical scaling laws with **short period waves (<1 s)**

SOURCE CHARACTERIZATION | ROCKFALLS

- LP surface wave inversion (T=40-150s) : Force
- Infer from Force : vitesse, acceleration, trajectory and mass

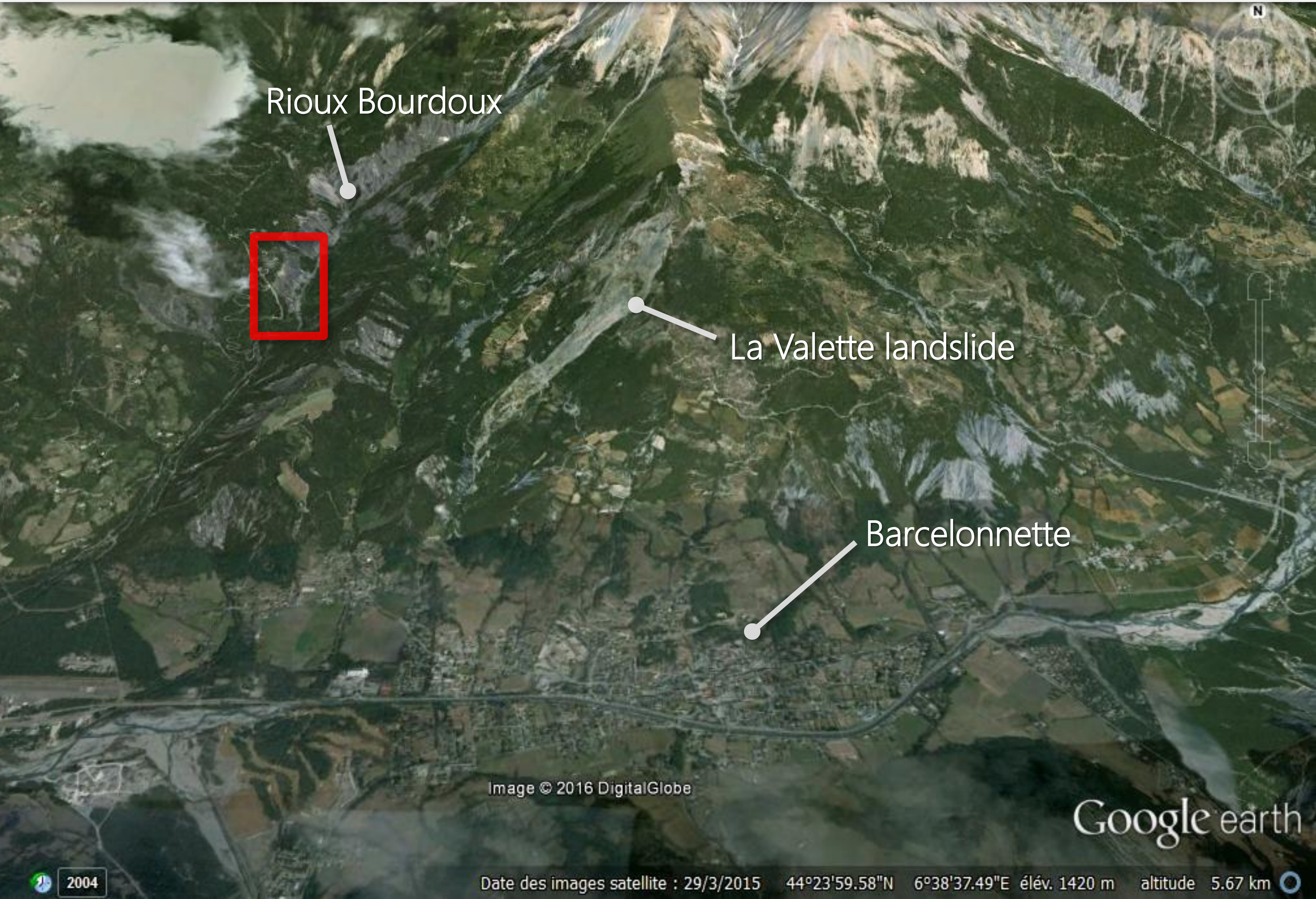




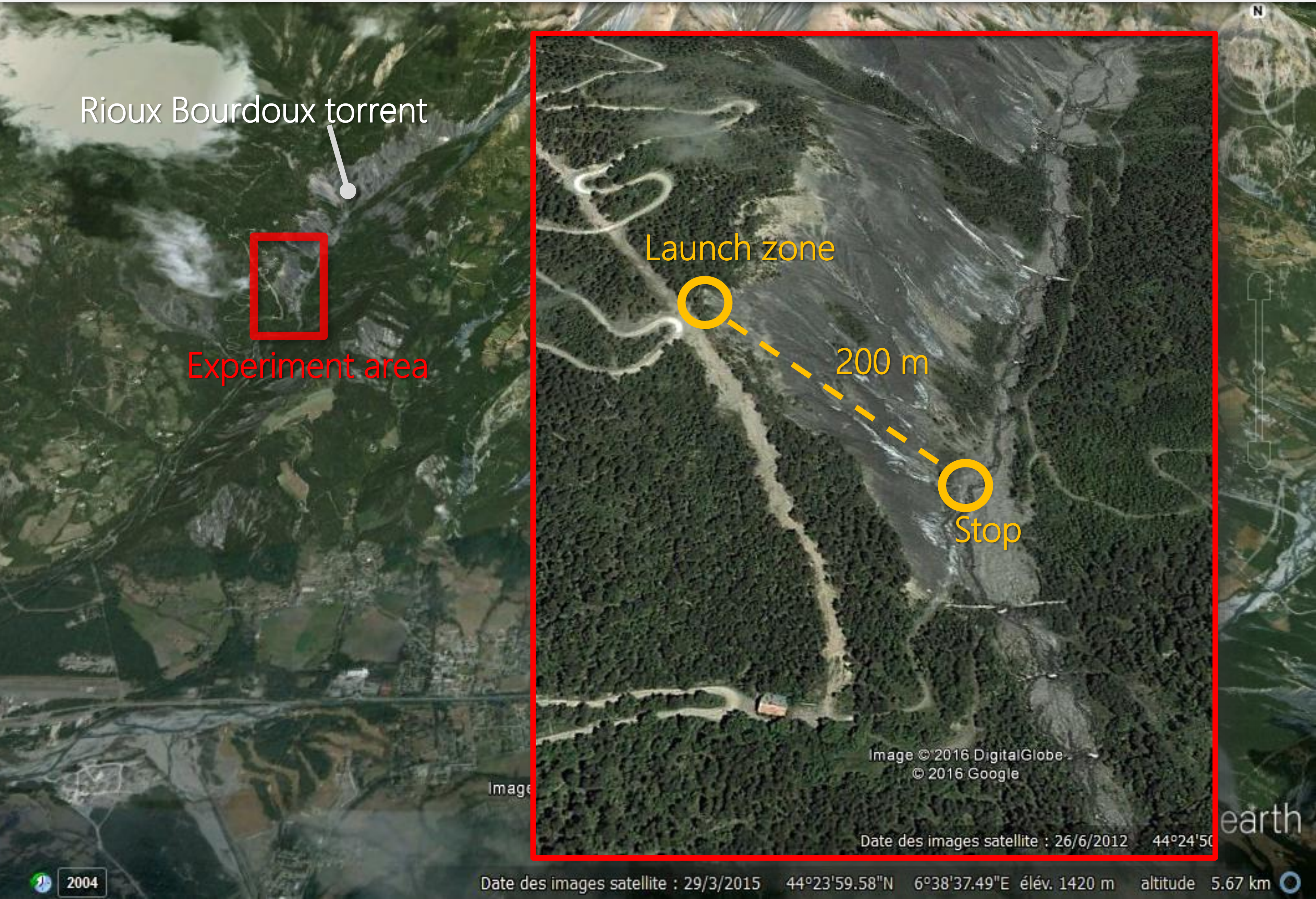


Limits : Only very large landslides
= <1% of events worldwide

SOURCE CHARACTERIZATION | **ROCKFALLS**



SOURCE CHARACTERIZATION | **ROCKFALLS**

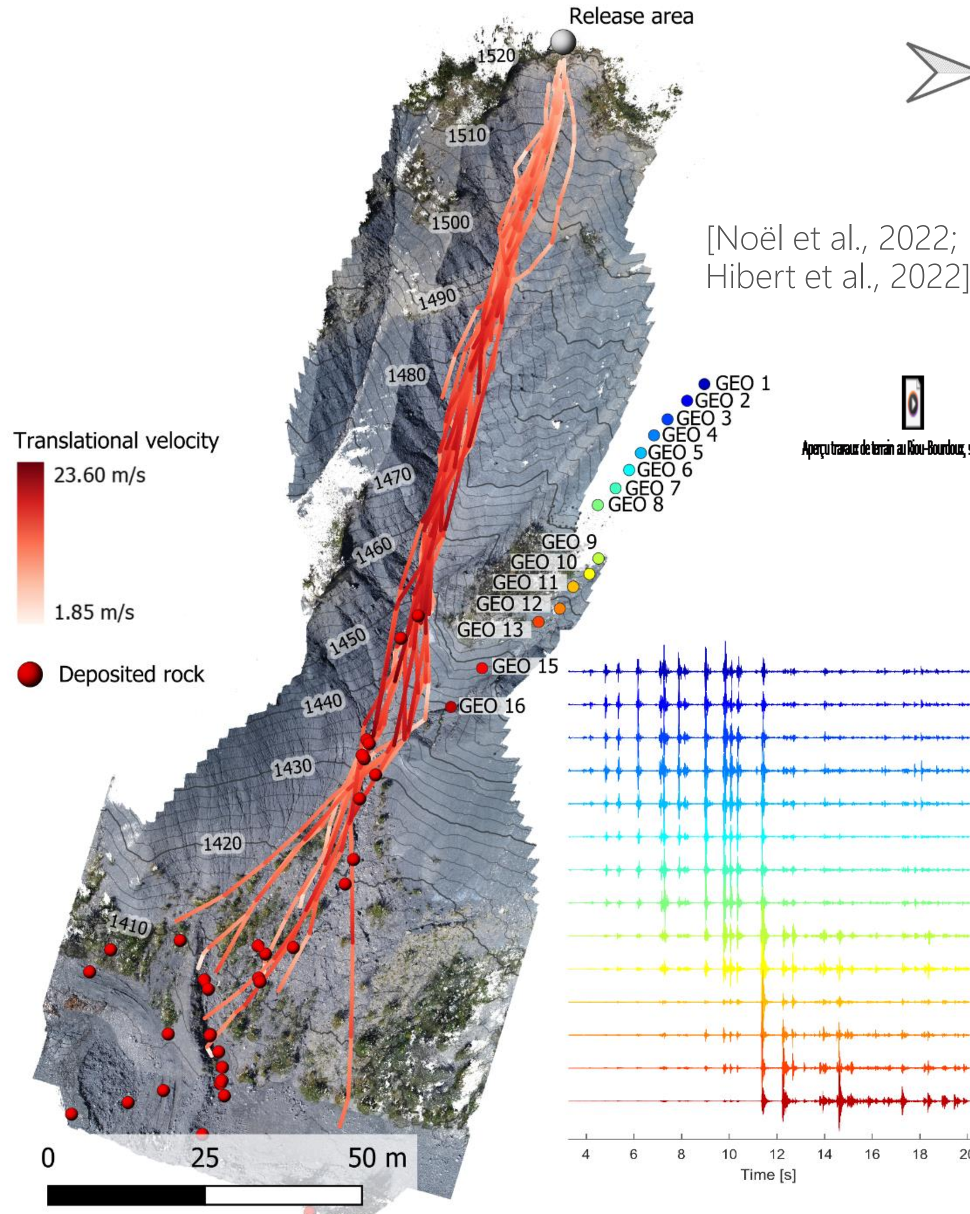


SOURCE CHARACTERIZATION | **ROCKFALLS**

Trajectory reconstruction :

Manual picking of the impact position and time

- ➢ Precize localisation thanks to DEM



[Noël et al., 2022; Hibert et al., 2022]

Aperçu travaux de terrain au Riou-Bourdoux, sept. 2018.mp4

From the trajectories :

Velocity, energies, momentum

$(mass \times velocity)$

SOURCE CHARACTERIZATION | **ROCKFALLS**

Machine learning prediction of the sources properties :

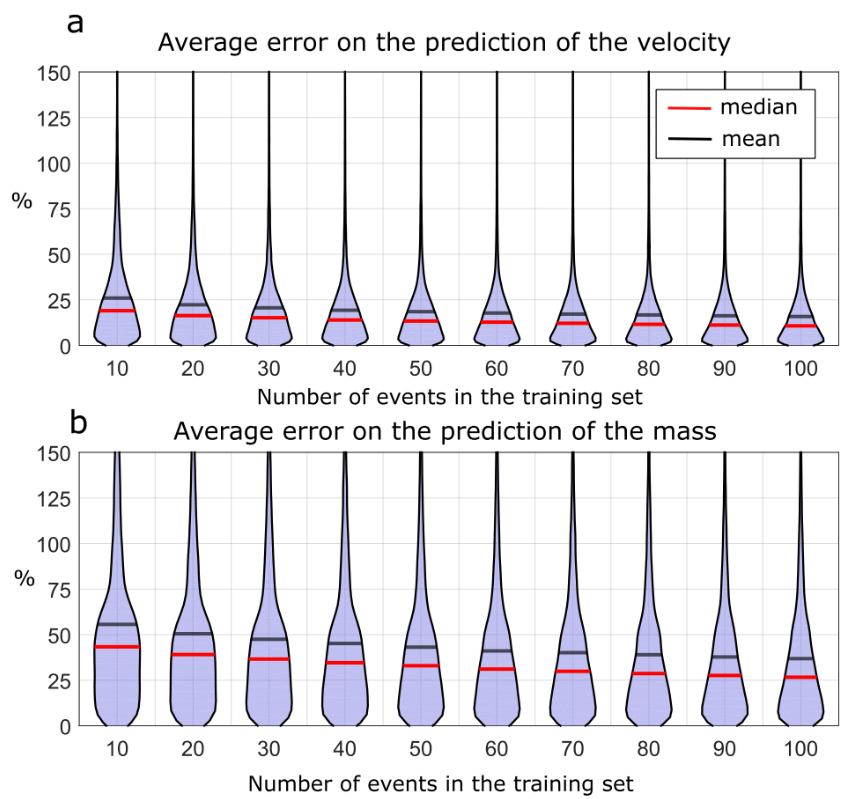
- Training and testing with features of 400 impacts signals
- Predictive model based on « Random Forests »
- Prediction of the mass and the velocity of the impactors

Results :

Median error on the **velocity** : 10%
Median error on the **mass** : 25%

✓ Lower uncertainties compared to physical scaling laws

✓ No need for the localization of the impact nor of a velocity model



[Noël et al., 2022; Hibert et al., 2022]