Self-supervised learning for (environmental) seismology

1ère Journées Epos France – Saint-Jean-Cap-Ferrat
10/11/2023

Clément Hibert | hibert@unistra.fr
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?*)
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)
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
**DETECTION | 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.]
Testing ensemble algorithms + curated features

Local scale:
- **Super-Sauze** [Provost et al., 2017] – 4 classes, ~900 eve. Success rate: 90%
- **Piton de la Fournaise** [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.]
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.
Dataset - Windowed catalogue

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

<table>
<thead>
<tr>
<th>Classes</th>
<th>Nb windows</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise (N)</td>
<td>220 798</td>
<td>17.11%</td>
</tr>
<tr>
<td>Ambient Noise (AN)</td>
<td>1 000 000</td>
<td>77.48%</td>
</tr>
<tr>
<td>Total Noise</td>
<td>1 220 798</td>
<td>94.59%</td>
</tr>
<tr>
<td>Rockfall (RF)</td>
<td>34 073</td>
<td>2.64%</td>
</tr>
<tr>
<td>Earthquake (EQ)</td>
<td>27 017</td>
<td>2.09%</td>
</tr>
<tr>
<td>Short Low Frequency (SLF)</td>
<td>5 456</td>
<td>0.42%</td>
</tr>
<tr>
<td>Microquake (MQ)</td>
<td>3 266</td>
<td>0.25%</td>
</tr>
<tr>
<td>Total Event</td>
<td>69 812</td>
<td>5.41%</td>
</tr>
<tr>
<td>Total</td>
<td>1 290 610</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

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

Confusion matrix:

<table>
<thead>
<tr>
<th></th>
<th>EQ</th>
<th>RF</th>
<th>MQ</th>
<th>AN + N</th>
<th>SLF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True Label</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EQ</strong></td>
<td>0.91</td>
<td>0.00</td>
<td>0.03</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>22387</td>
<td>115</td>
<td>727</td>
<td>205</td>
<td>1081</td>
</tr>
<tr>
<td><strong>RF</strong></td>
<td>0.01</td>
<td>0.93</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>262</td>
<td>29449</td>
<td>442</td>
<td>788</td>
<td>631</td>
</tr>
<tr>
<td><strong>MQ</strong></td>
<td>0.00</td>
<td>0.01</td>
<td>0.92</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>9</td>
<td>704</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td><strong>AN + N</strong></td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>502</td>
<td>6532</td>
<td>1335</td>
<td>1202438</td>
<td>4988</td>
</tr>
<tr>
<td><strong>SLF</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.12</td>
<td>0.04</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>9</td>
<td>350</td>
<td>108</td>
<td>2479</td>
</tr>
</tbody>
</table>

Precision: 0.97
Recall: 0.91
F1-score: 0.94

Precision: 0.82
Recall: 0.93
F1-score: 0.87

Precision: 0.2
Recall: 0.92
F1-score: 0.33

Precision: 1.0
Recall: 0.99
F1-score: 0.99

Precision: 0.27
Recall: 0.84
F1-score: 0.41
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 process *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], …
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

Mayotte volcano - REVOSIMA catalogue
- 2 stations: IF07C & IF07D
- From 1/10/19 to 19/11/19

<table>
<thead>
<tr>
<th>Classes</th>
<th>Nbr Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volcano-Tectonic earthquakes (VT)</td>
<td>2,008</td>
</tr>
<tr>
<td>Hydro-Acoustic signals (HA)</td>
<td>1,626</td>
</tr>
</tbody>
</table>

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

Data transformation

HD images

256 x 256
Self-supervised Learning

Mayotte volcano - REVOSIMA catalogue
- 2 stations: IF07C & IF07D
- From 1/10/19 to 19/11/19

SimSiam = 512D
Self-supervised Learning

Mayotte volcano - REVOSIMA catalogue
- 2 stations: IF07C & IF07D
- From 1/10/19 to 19/11/19

\[ \text{SimSiam} = 512D \]

\[ \text{t-SNE} = 2D \]

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

Mayotte volcano - REVOSIMA catalogue:
- 2 stations: IF07C & IF07D
- From 1/10/19 to 19/11/19

SimSiam = 512D

t-SNE = 2D

DBSCAN = Clusters

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

Mayotte volcano - REVOSIMA catalogue
- 2 stations: IF07C & IF07D
- From 1/10/19 to 19/11/19
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
Thank you!

Contact: hibert@unistra.fr
DETECTION | CLASSIFICATION

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

> Need for comprehensive catalogues of landslides
Training Set : 2 classes

Earthquakes :

- 290 Earthquakes recorded by the Alaskan 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
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

Grout 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.]
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.
Training set: 2 classes

**Glacial Earthquake**

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

**Earthquakes**

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

**GEQ**

- **444** GEQ (M > 4.5)
- **3424** signals
- Known events (GCMT project, Ekström et al.)
Application to the GLISM network on 844 days

- 5791 events > 1670 new GEQ confirmed manually
  = 4x the GCMT Cat.
- Events discarded: 758 EQ, possible + GEQ but with signal only on one station

<table>
<thead>
<tr>
<th>Training percent</th>
<th>GEO</th>
<th>EQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 5.0%</td>
<td>2927 +/-56</td>
<td>325 +/-56</td>
</tr>
<tr>
<td>b. 10.0%</td>
<td>2797 +/-34</td>
<td>284 +/-34</td>
</tr>
<tr>
<td>c. 25.0%</td>
<td>2376 +/-19</td>
<td>191 +/-19</td>
</tr>
<tr>
<td>d. 50.0%</td>
<td>1592 +/-11</td>
<td>119 +/-11</td>
</tr>
</tbody>
</table>

Classes: Pirot et al.
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
BARCELONNETTE

La Valette

Landslide

Rioux Bourdoux torrent

Experiment area

Launch zone

Stop

200 m

SOURCE CHARACTERIZATION | ROCKFALLS
Trajectory reconstruction:
Manual picking of the impact position and time
➢ Precise localisation thanks to DEM

From the trajectories:
Velocity, energies, momentum \((mass \times velocity)\)
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]