ULB UNIVERSITÉ LIBRE DE BRUXELLES

MLG MACHINE LEARNING GROUP

Thesis update - 6/11/2023

Pascal Tribel

Faculté des Sciences
Université Libre de Bruxelles

ULB Faculté des Sciences

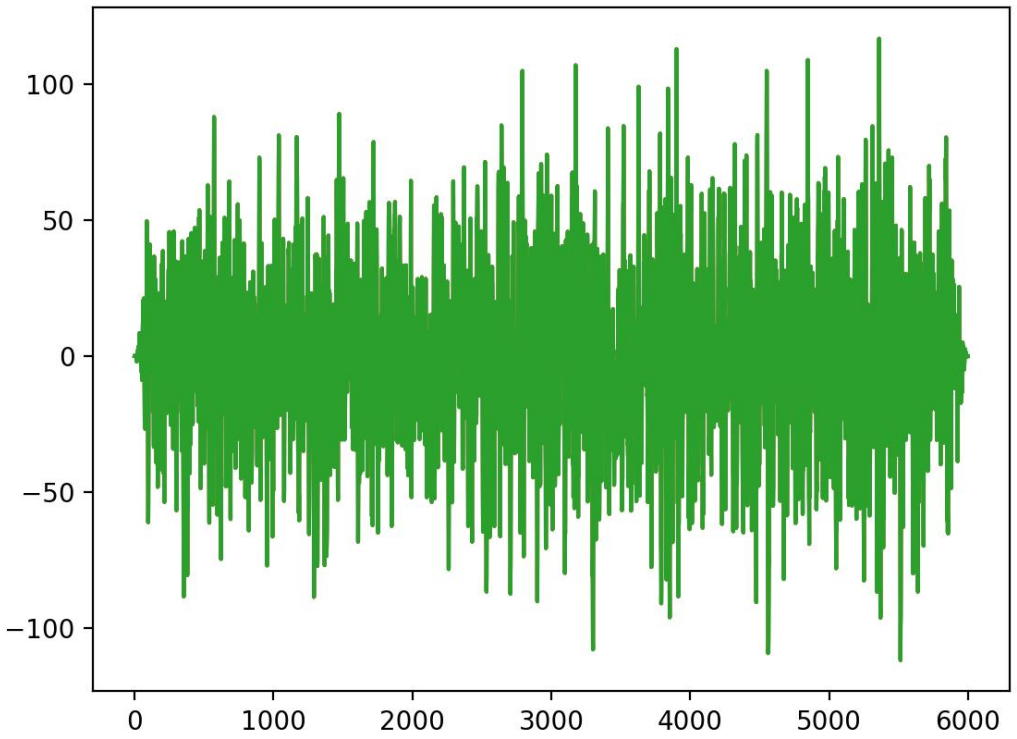
Outline

Where I am

- Two kinds of data:
  - Sismographs: STandford EArthquakes Dataset
  - Distributed Acoustic Sensing: Laacherzee (tdms), LaPalma (bin)
- The problems depend on the data type
  - Sismographs
    - More advanced problems
    - Accurately labelled data
    - Worse resolution
  - DAS
    - Basic problems
    - Great resolution, really massive data
    - Almost unlabelled data

Outline

STEAD



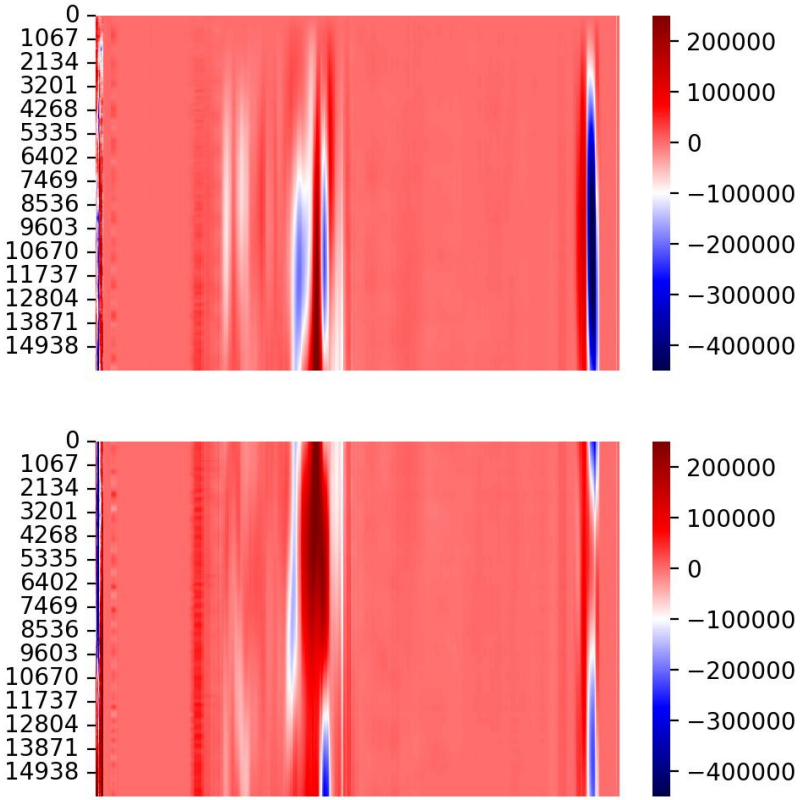
1 minute of STEAD data

STEAD

- 1-minute long data, 90Gb, labeled
- SOTA: P and S arrivals, earthquake labelling, magnitude prediction
- Ideas: time series forecasting, data evolution anticipation
- Quantity is not a problem: 90Gb of data, great labels
- Main library: Seisbench
- Tested with a NN to predict the magnitude of earthquakes for each 1-minute sequence

Outline

DAS



4 seconds of Laacherzee DAS data

DAS

- SOTA: data collection, tremor detection
- Challenges: long, massive, and noisy data. Even displaying the data is hard. Streaming techniques are mandatory.
- Incomplete data labels: no indication of earthquake presence/magnitude... but location/time is present
- Ideas: put in relation with labeled seismograph data to infer the labels
- No great standard library to manage the data (despite lightguide/pyrocko), no great streaming workflow library