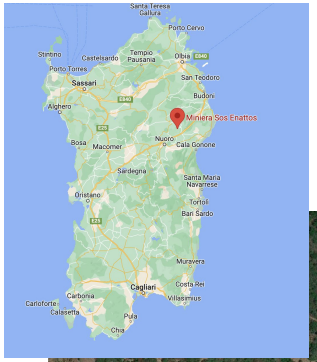
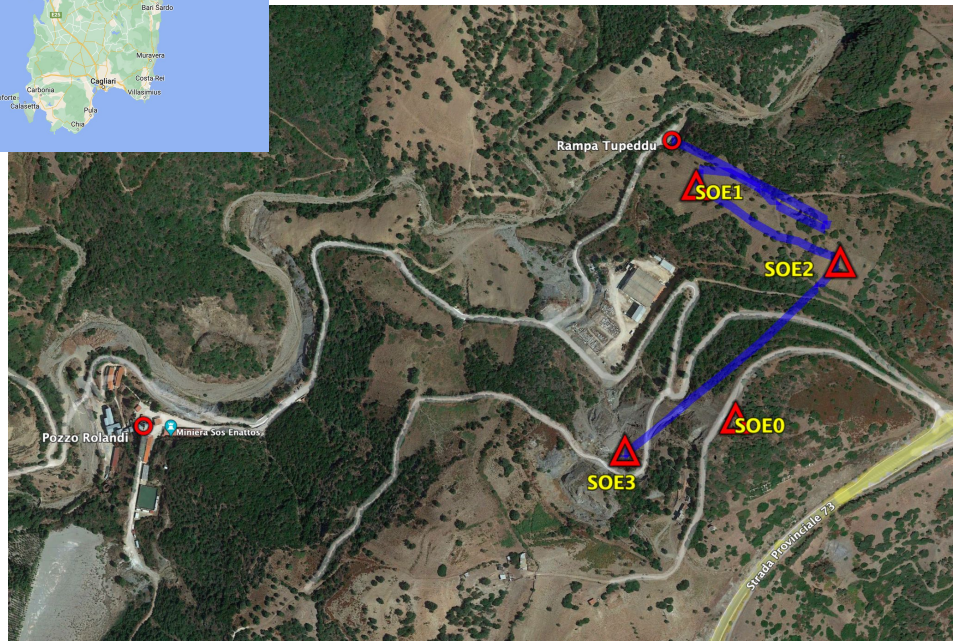
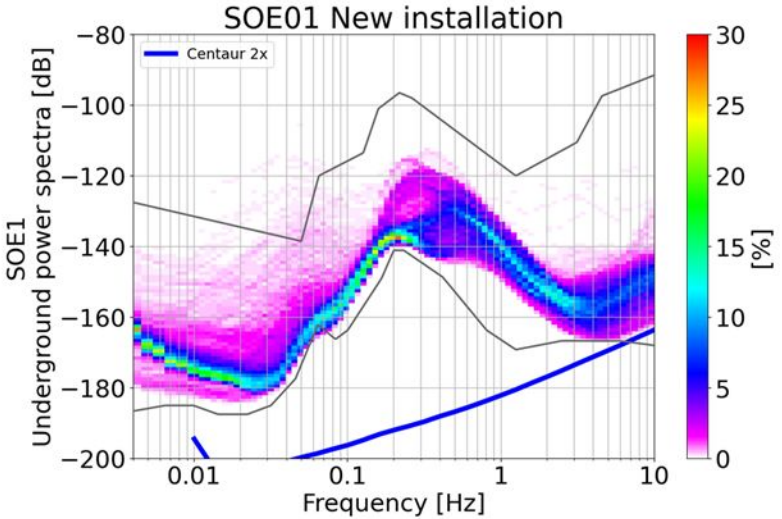
Automatic Hierarchical Exploration of Continuous Seismic Signal with Unsupervised Learning

Christopher Zerafa, Carlo Giunchi, Leonard Seydoux
SOME group, INGV, Pisa
christopher.zerafa@ingv.it

Motivation

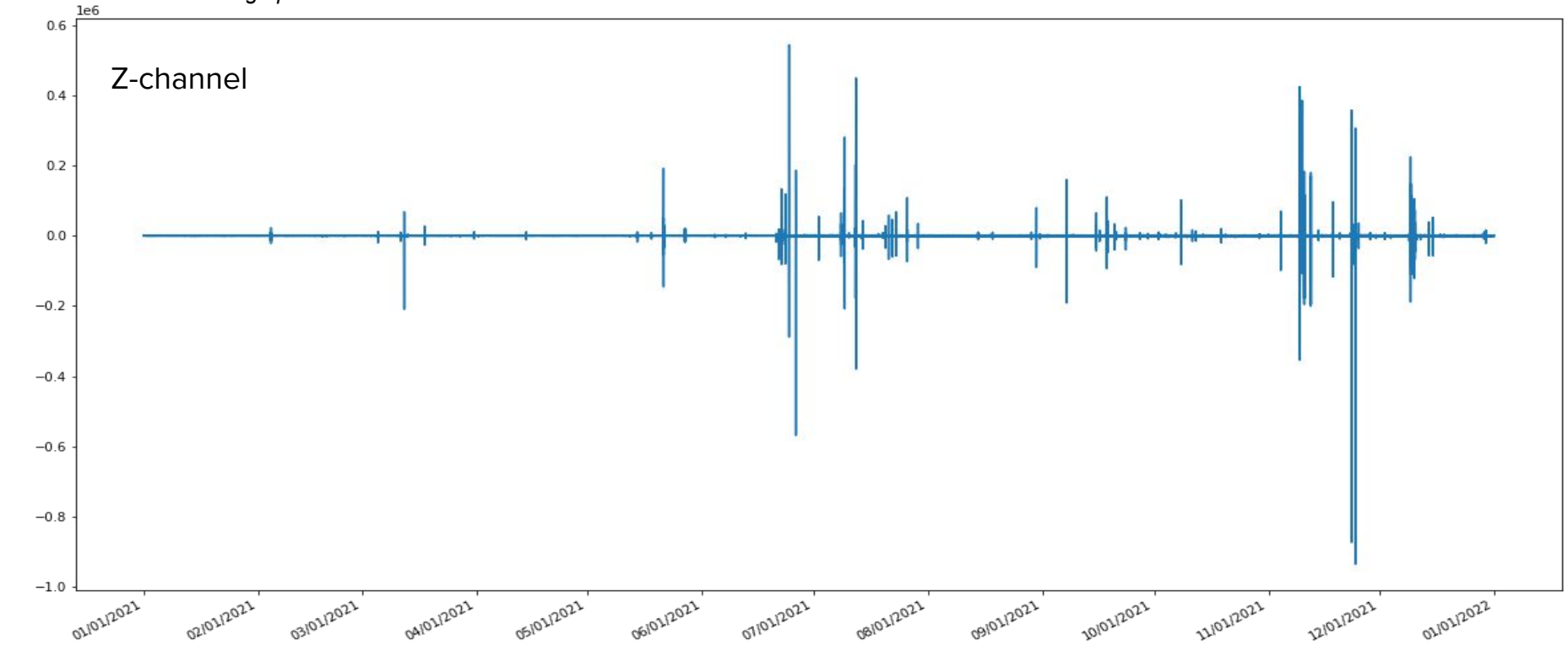

Miniera Sos Enattos
Sardegna
Nuoro
Cagliari
Sassari
Alghero
Oristano
Porto Torres
Olbia
Tempio Pausania
Castelsardo
Bosa
Macomer
Cala Gonone
Santa Maria Navarrese
Tortolì
Bari Sardo
Muravera
Costa Rei
Villasimius
Carbonia
Carloforte
Calasetta
Pula
Porto Cervo
San Teodoro
Budoni
Stintino
Santa Teresa Gallura


Rampa Tupeddu
SOE1
SOE2
SOE0
SOE3
Pozzo Rolandi
Miniera Sos Enattos
Strada Provinciale 73


SOE01 New installation
Centaur 2x
SOE1
Underground power spectra [dB]
−80
−100
−120
−140
−160
−180
−200
0.01
0.1
1
10
Frequency [Hz]
30
25
20
15
10
5
0
[%]

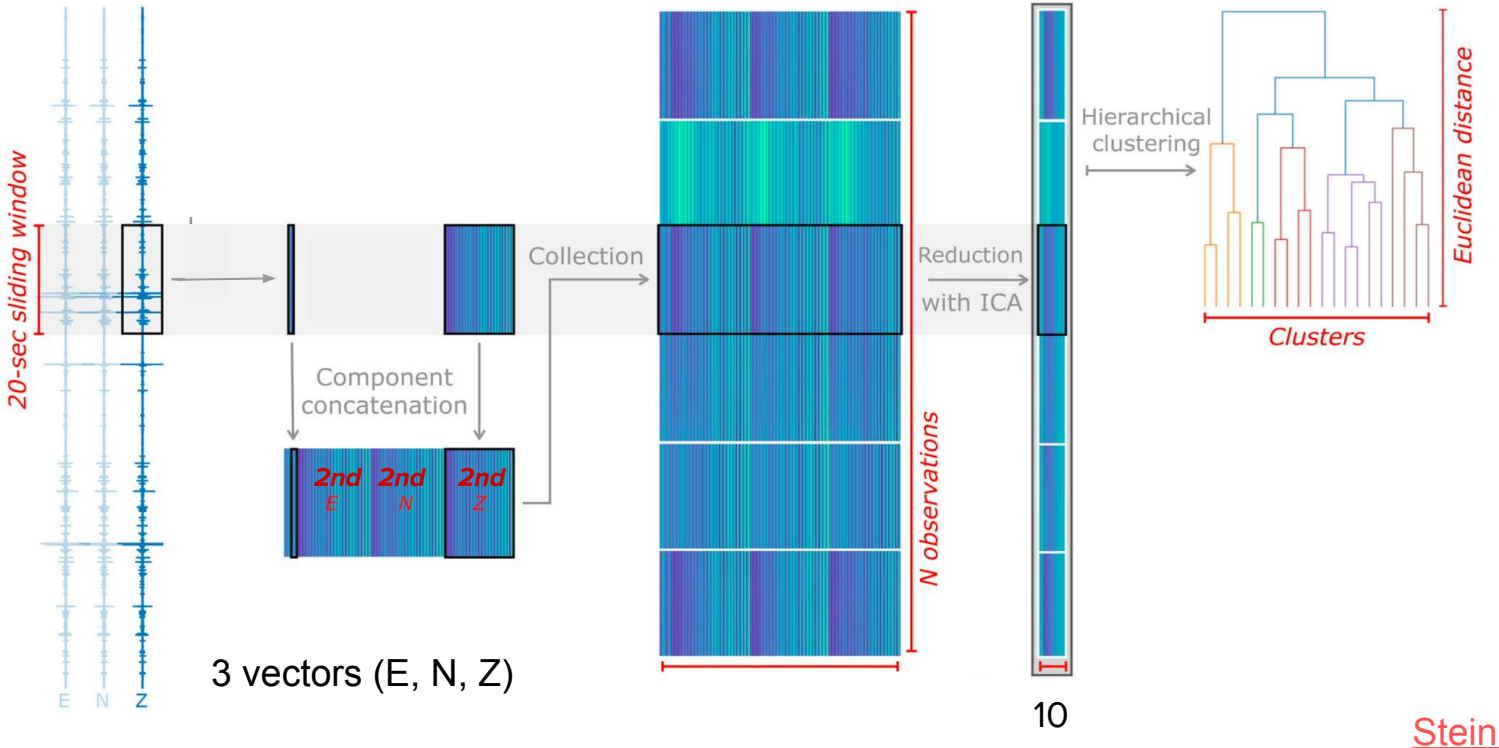
Data

- *linear detrend*
- *merge interpolation*
- *decimation to 50Hz*
- *1Hz high pass*



Previous Work

1. Continuous Seismograms
2. Deep Scattering Spectrum
3. Reduction
4. Clustering
5. Manual analysis of clusters



Steinmann et al. 2022

Why not just apply previous methodology?

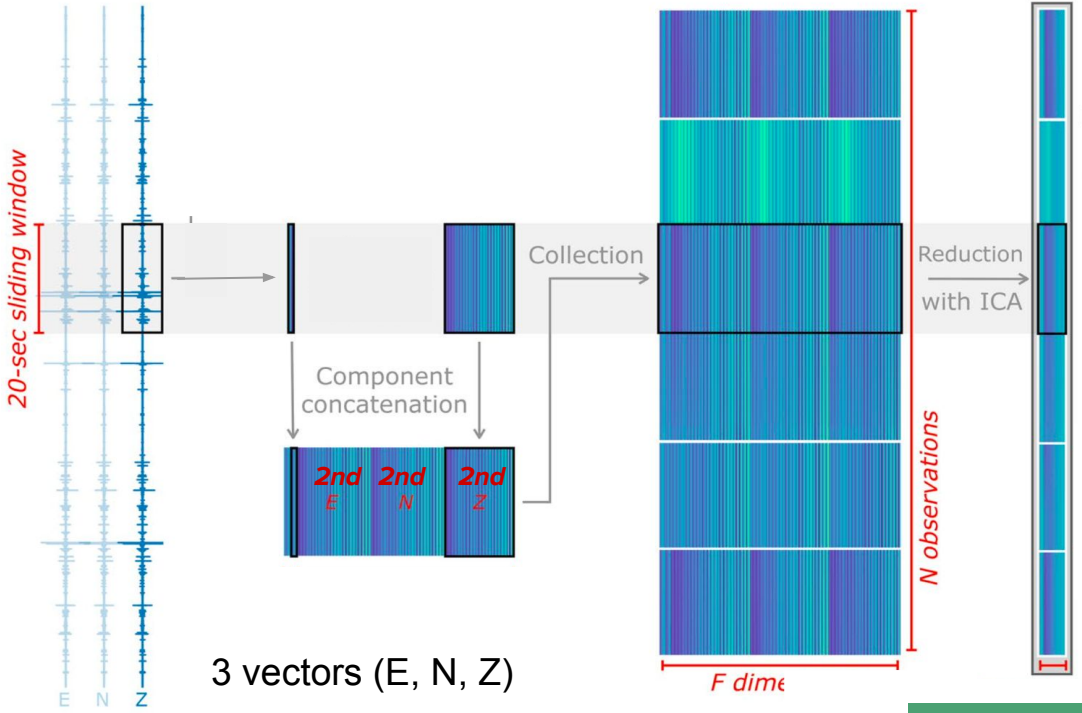
What challenges are we facing?

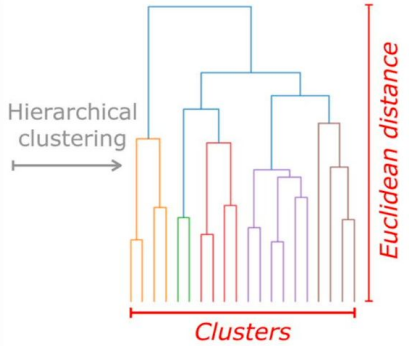
- Extremely quite
- Large data (3+ years)

Our Approach

Modification

1. Continuous Seismograms
2. Deep Scattering Spectrum
3. Reduction
4. Clustering
5. Manual analysis of clusters





Modification

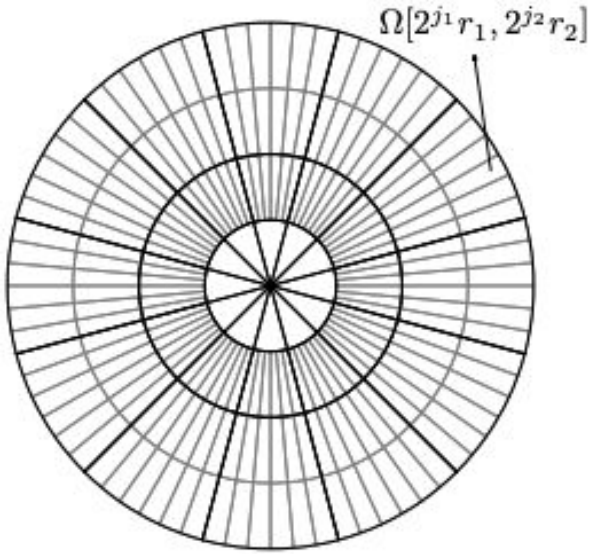
Modification

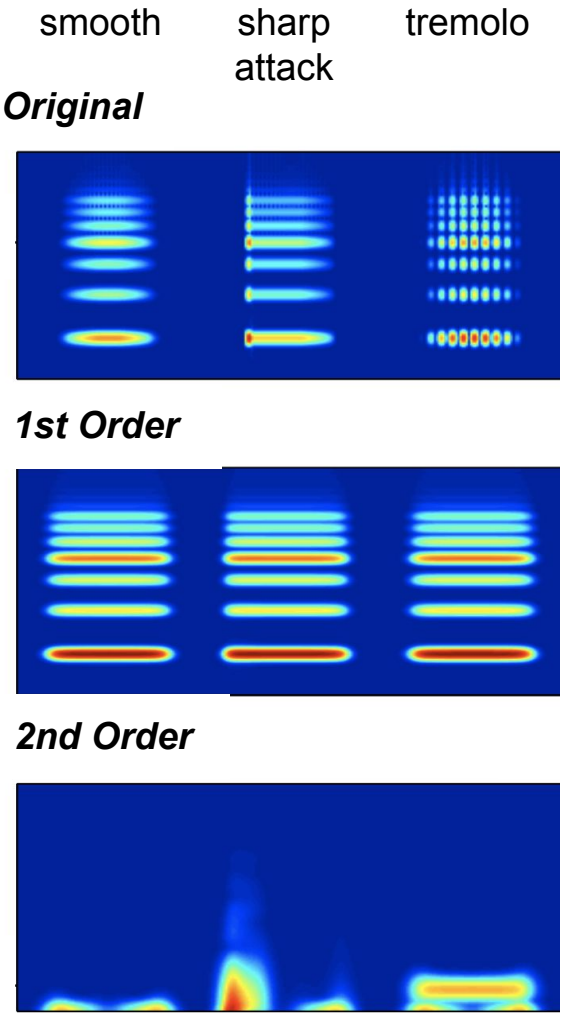
Steinmann et al. 2022

Deep Scattering Spectrogram

Deep Scattering Spectrum - Andén & Mallat 2011
(https://www.di.ens.fr/data/publications/papers/1304.6763v1.pdf)

- Locally translation invariant
- Stable to time-warping deformations
- Extension of Mel-frequency cepstral coefficients via cascades of wavelet convolutions and modulus

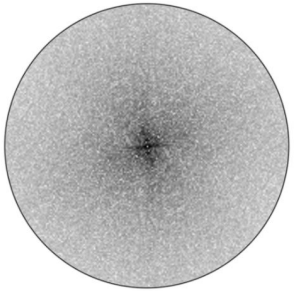


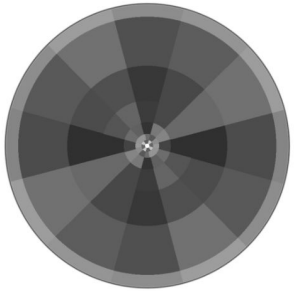


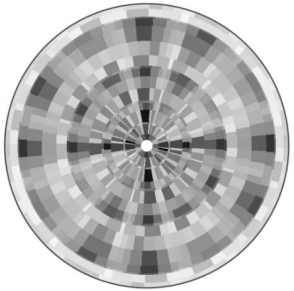
Deep Scattering Spectrogram (cont.d)

Why?

- Introduces sparsity
- Is the additional processing step worth it?

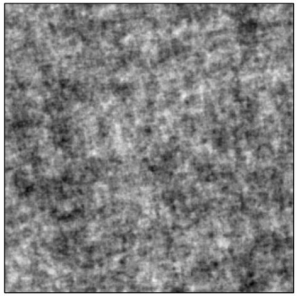


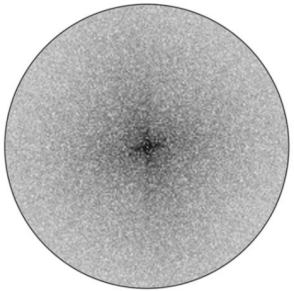




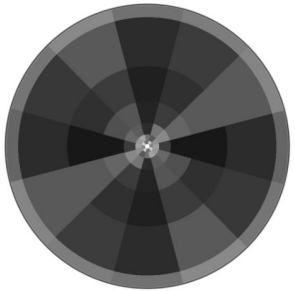
How?

Kymatio - Andreux et al. 2020 (https://jmlr.org/papers/v21/19-047.html)

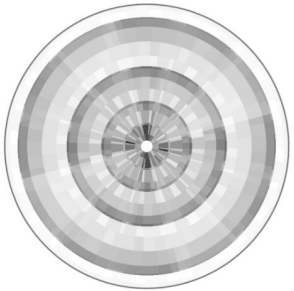




Same Power Spectrum

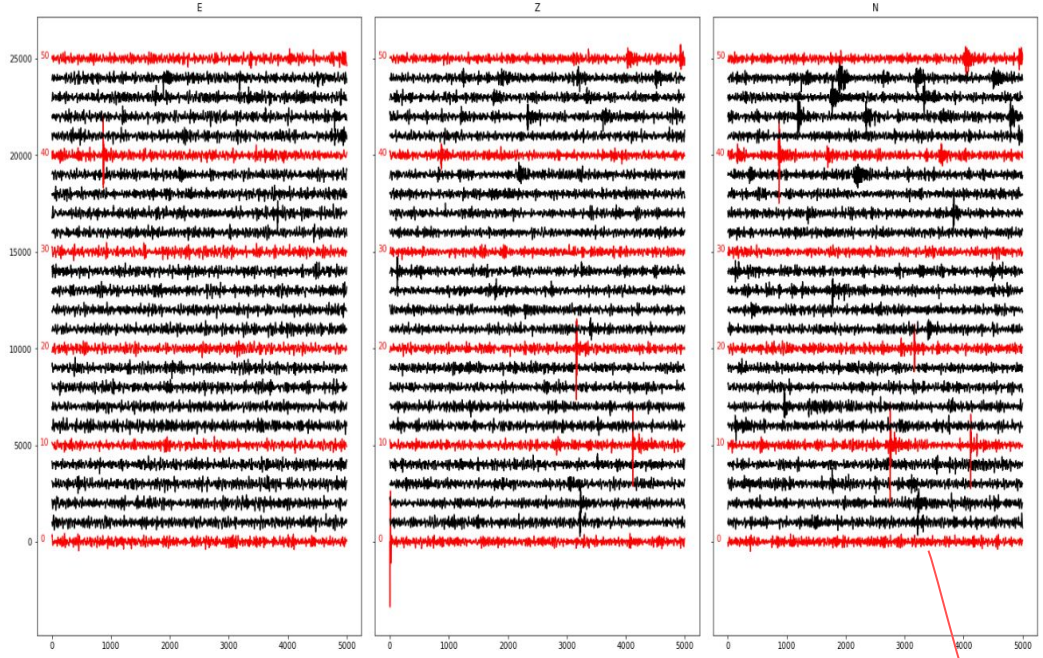


Nearly same 1st Order

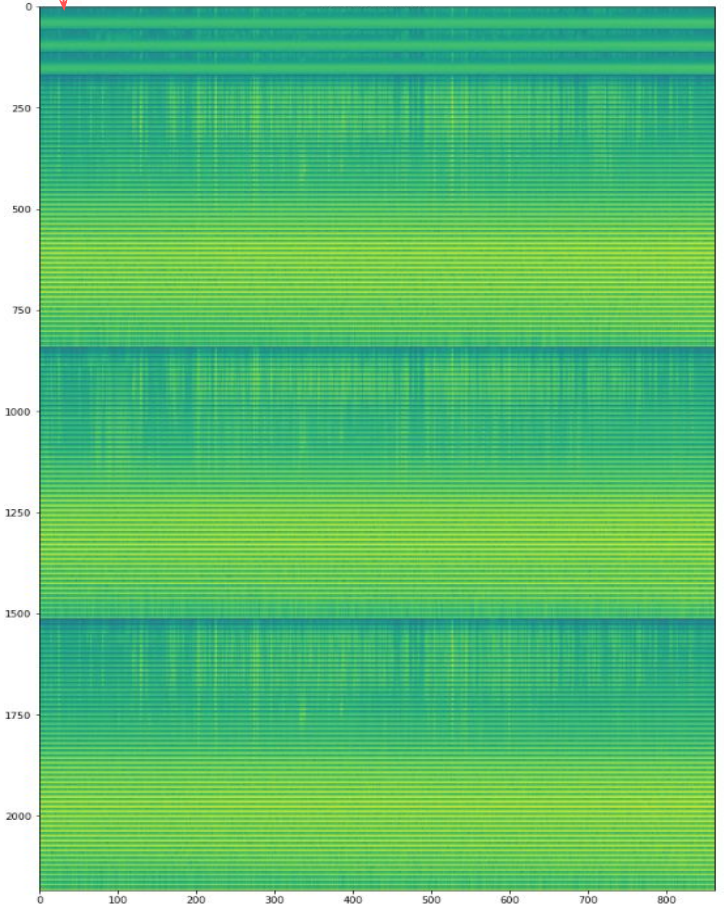


Very different 2nd Order

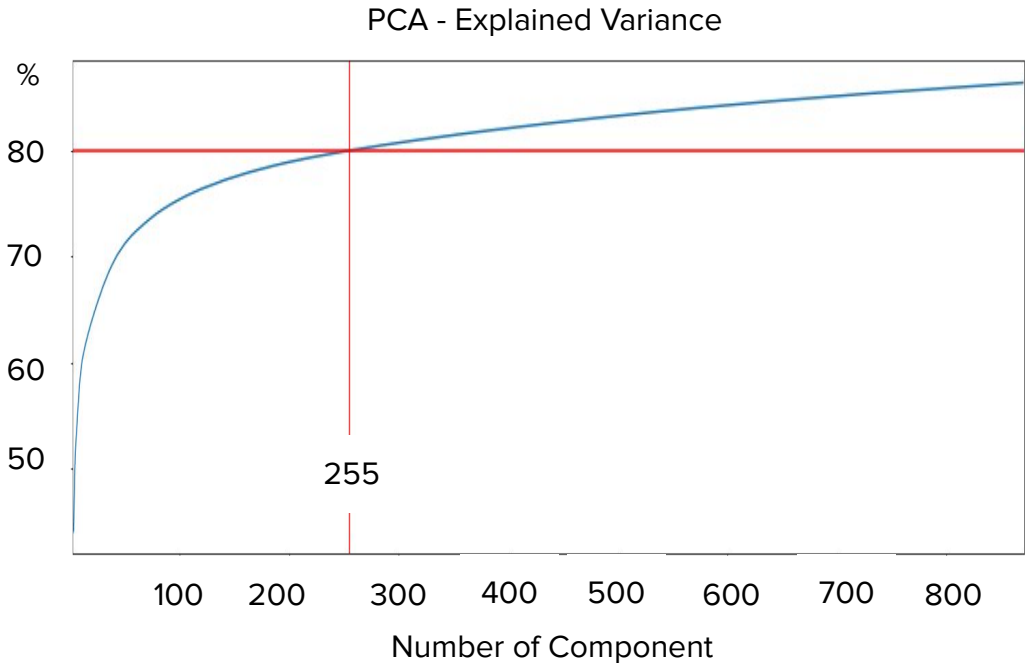
Deep Scattering Spectrum

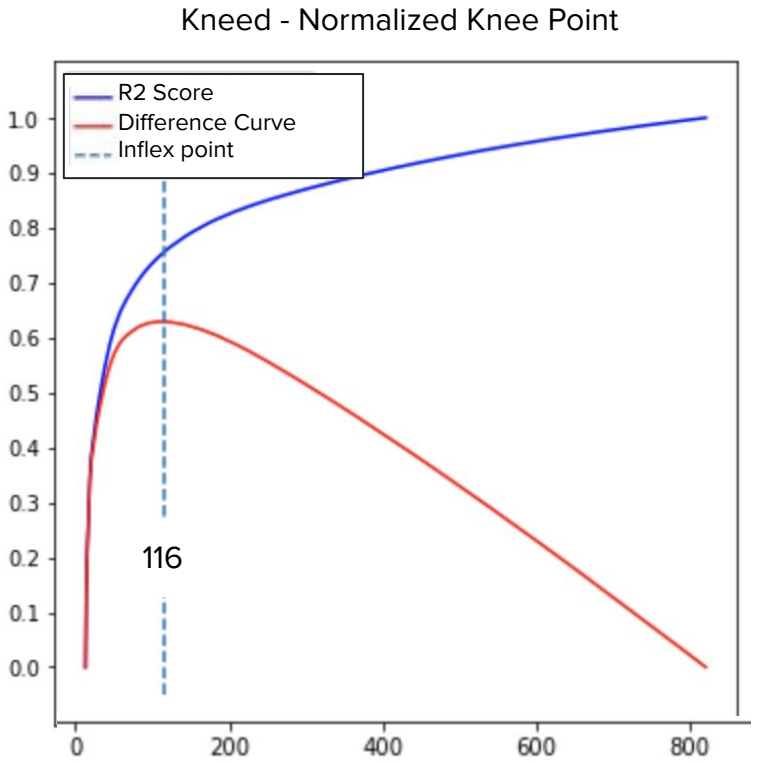


Large Dataset
314,630-104,091

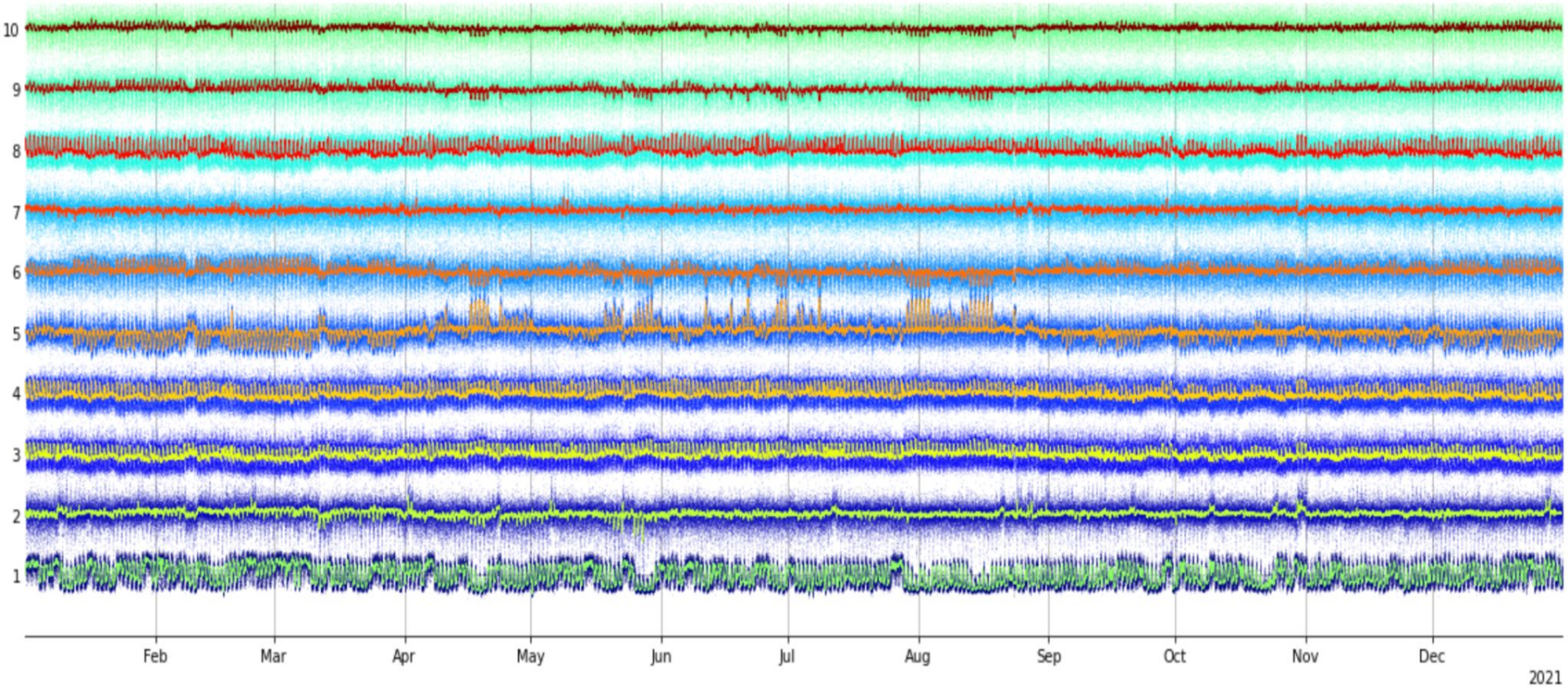


Reduction - PCA or Kneed

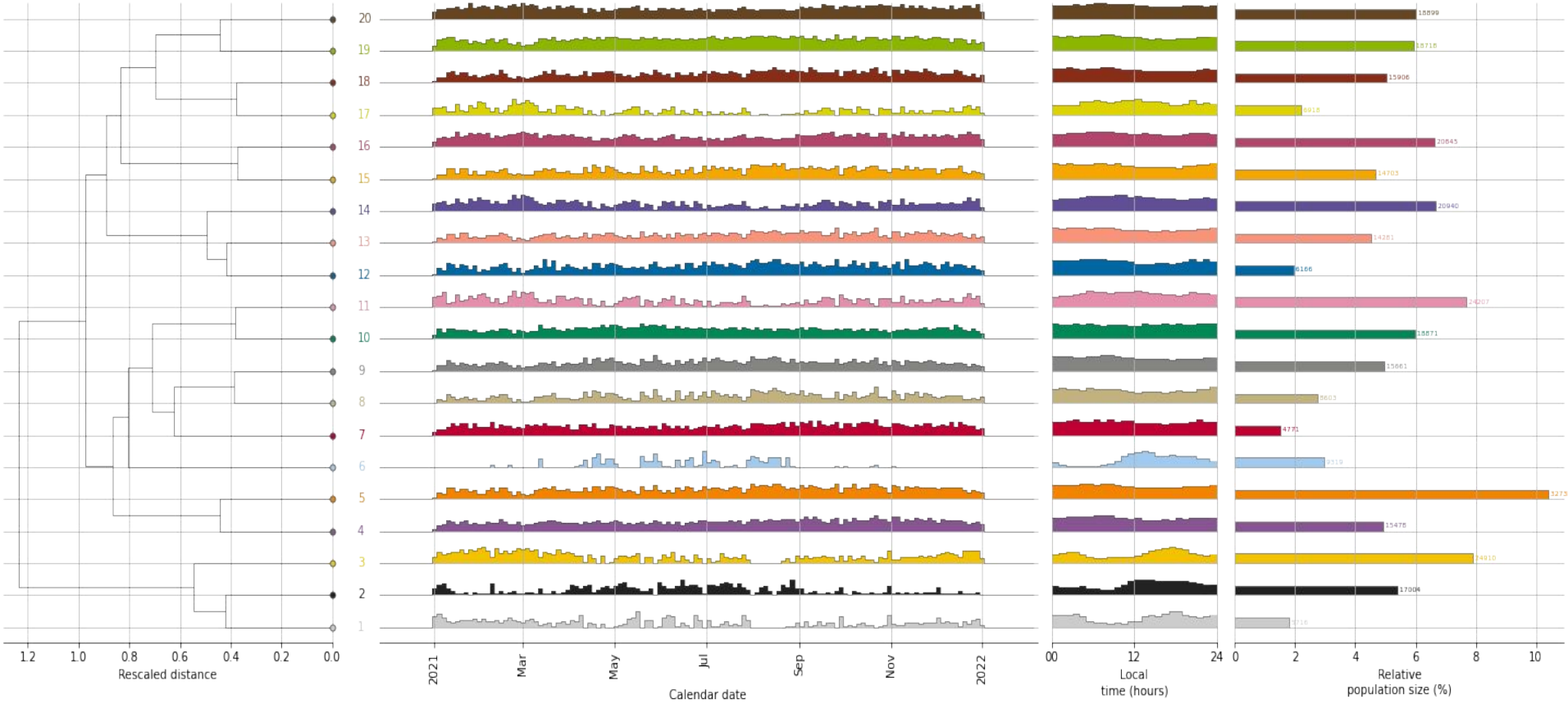




Source Separation - ICA



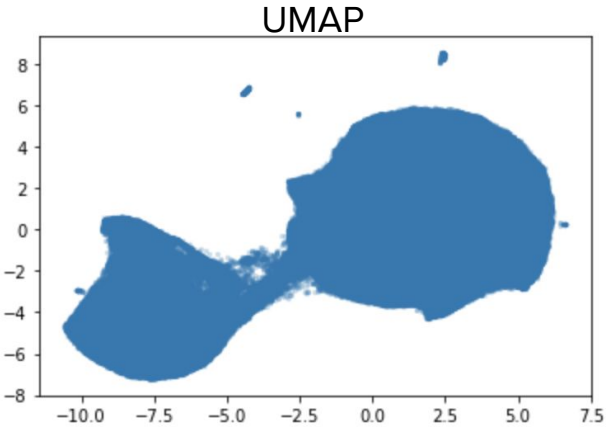
Clustering


20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Rescaled distance
2021
Mar
May
Jul
Sep
Nov
2022
Calendar date
00
12
24
Local time (hours)
0
2
4
6
8
10
Relative population size (%)
18899
15718
15906
8918
20845
14703
20940
14281
6166
24207
18871
15661
5603
4771
2519
32736
15478
19910
17004
3716

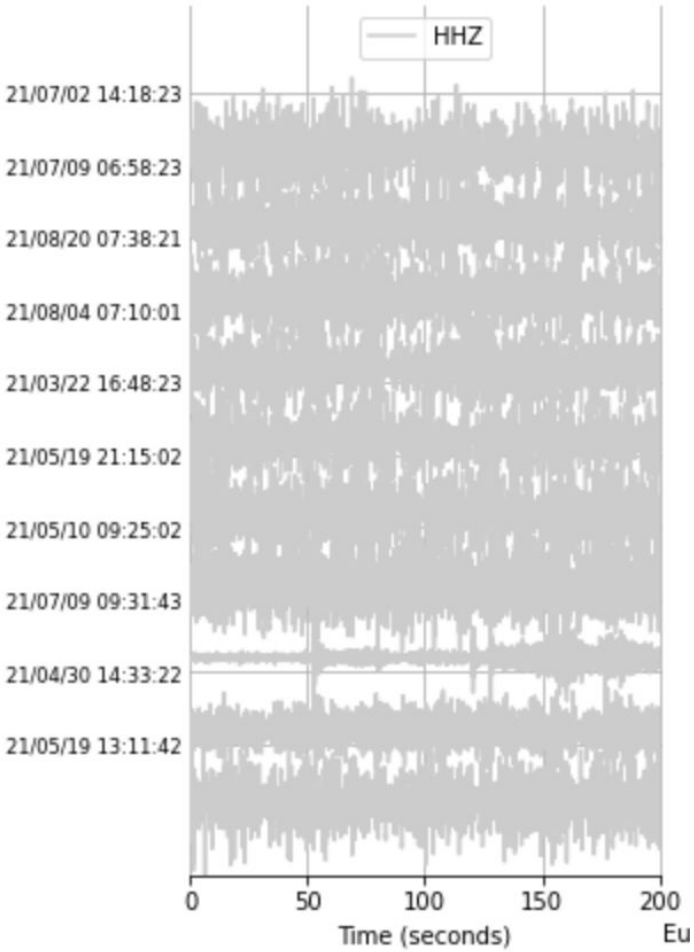
Interpretation of the clusters

Still a works in progress

Some signal has been identified

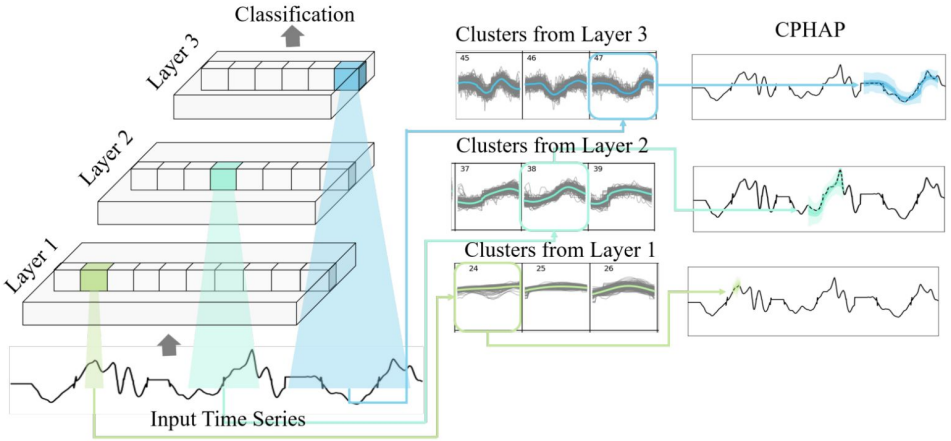


Cluster 6



Way Forward

- Identify the optimal approach for the reduction transform
  - AutoEncoders to achieve 100% compression
- Non-parametric unsupervised characterization of seismogram clusters. (Rojat 2021)



"***Gradient*Input***" technique used to highlight areas of a seismogram that contributed towards making a particular prediction.

What about g2net?

2020 g2net STSM *alumni*

We can run this workflow on

- seismic data from the EGO site,
- classify any seismic signal,
- calibrate interferometer prior to the next run

STSM anyone?

Thank you

Questions?