# Unsupervised deep learning method to automate the search of seismic signals

**Sarah MOUAOUED**
**sarah.mouaoued@univ-grenoble-alpes.fr**
**Michel CAMPILLO and Léonard SEYDOUX**

## Context

On the 4th and 6th of July 2019, at 17:33 TU and 03:19 TU respectively 2 earthquakes of magnitude 6.4 and 7.1 occurred in Ridgecrest, California. The fault system is defined by 2 faults perpendicular to each other (Figure 1). This is an area well covered by seismological network stations and seismologists didn't detect any precursor events. But is it because we are not able to detect them? With developments in artificial intelligence and data classification, it becomes possible to automatically separate noise data based on the physical origins of those signals. An unsupervised deep learning method has been developed in our group, this will allow a more detailed study of the seismic data.

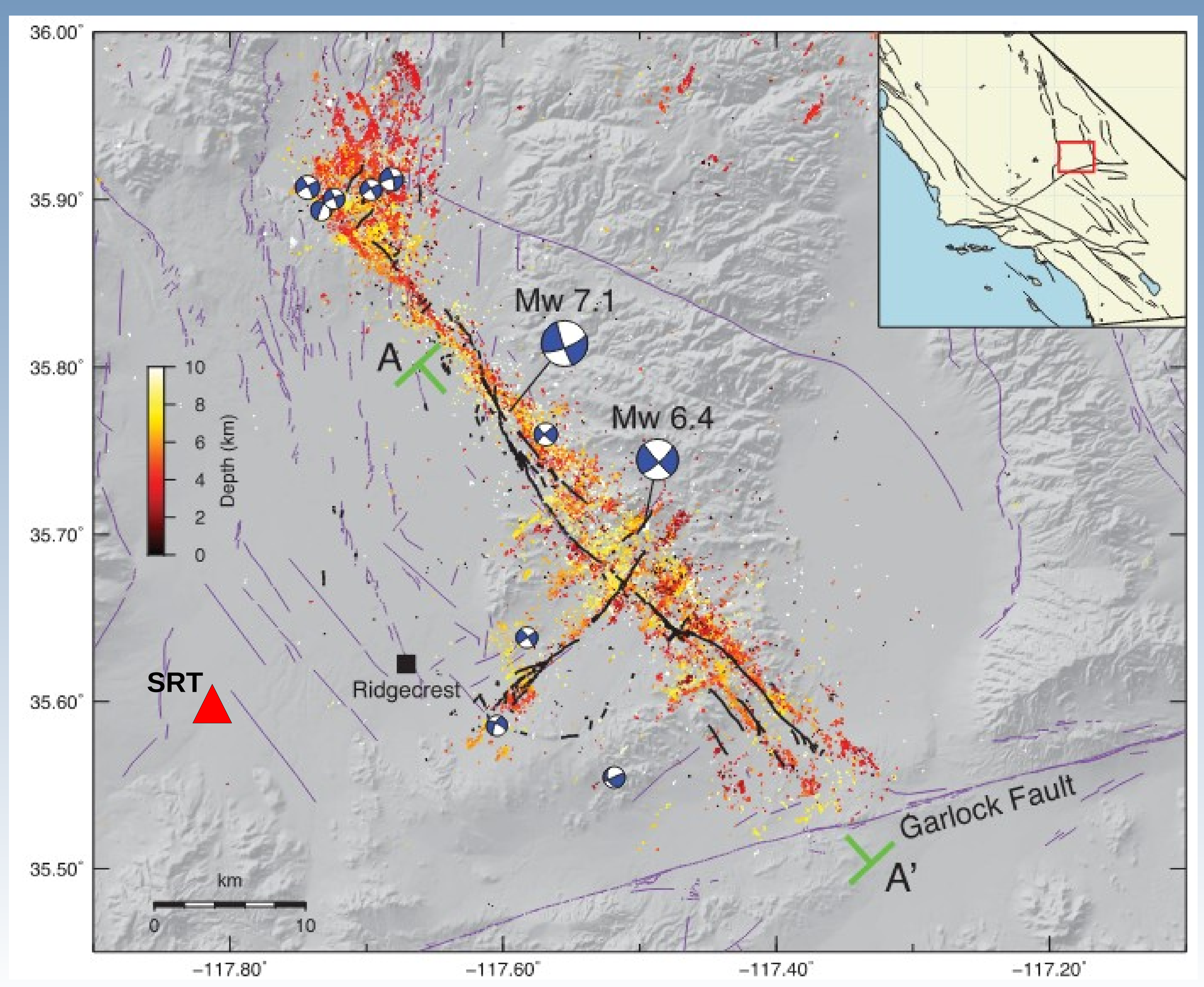



**Figure 1.** Map view of Ridgecrest seismicity (Ross et al., 2019).

## Method

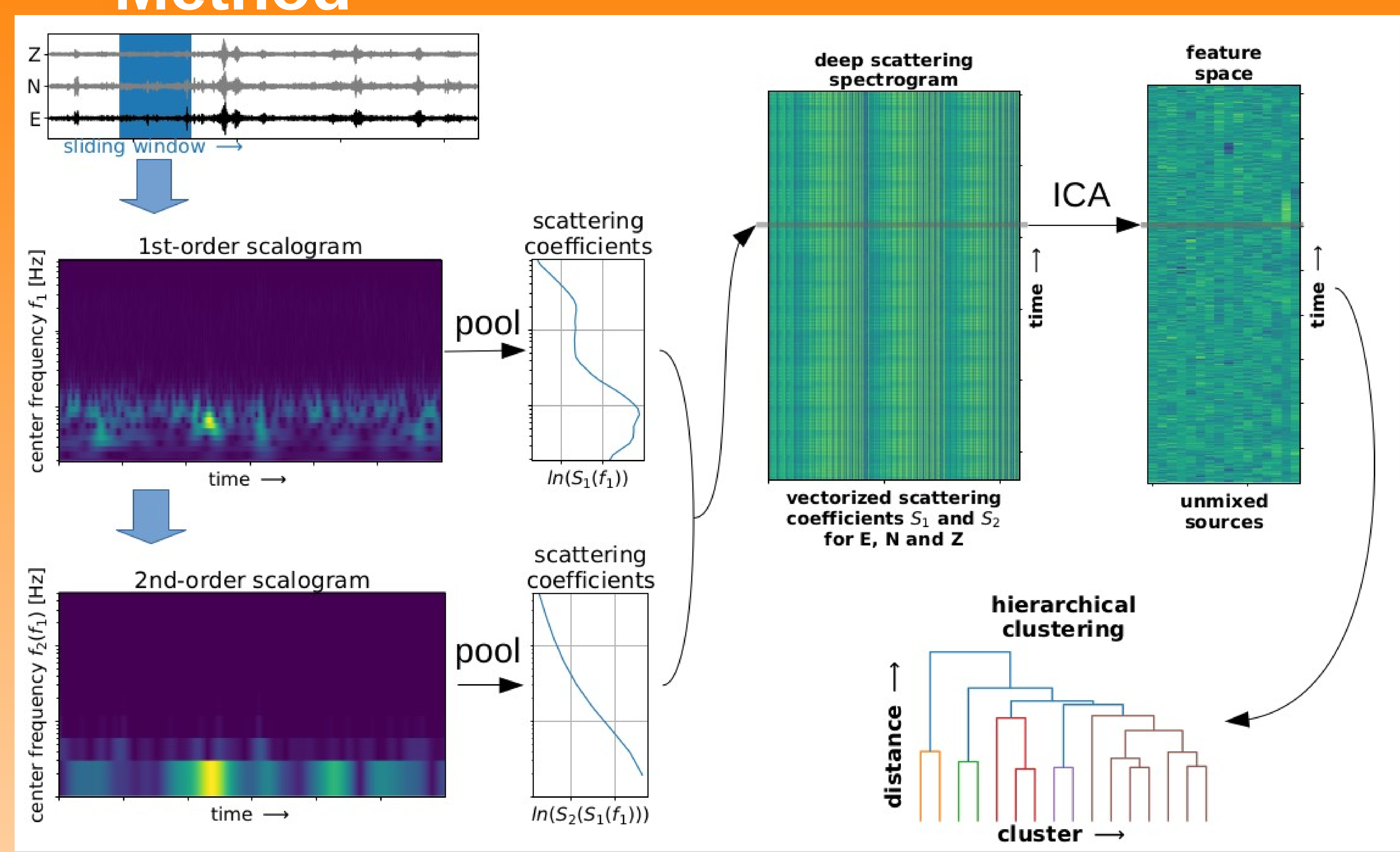



**Figure 2.** Sketch of the hierarchical clustering (Steinmann et al. ,2022)

To classify seismic data we used hierarchical clustering to compute a dendrogram (bottom right of Figure 2). The distance between clusters (which is a mix between an euclidean distance and a ward normalization) is measured according to features these ones being defined thanks to a scattering network. The scattering network is defined as a neural network where filters are not learned but are defined as Morlet wavelets. It allows a cascade of time-frequency analyses thanks to the wavelet's convolution of time windows. This process has been introduced by Mallat (2014) which contains more information than a Fourier decomposition. Then the ICA (Independent Composant Analysis) allows defining the features space by reducing the number of dimensions.

### Hierarchical clustering for 304 days of data from SRT station

We clustered **304 days** of continuous data from the 1st of January to the 1st of November, from the SRT station located west of the fault system (Figure 1) with a 40 Hz sampling frequency.

We segmentized the waveforms into **600 sec long windows** and filtered with a high-pass at 0,5 Hz to avoid microseisms. For the scattering transform, we choose for the first layer 6 octaves with a resolution of 8 wavelets for each and for the second layer 10 octaves with a resolution of 2 wavelets for each.

The features space is reduced to **20 dimensions** using Independ Component Analysis (Figure 3). We can see that some axes seem to correspond to noise (9, 10 ..) and from the beginning of July some of them seem to be characterized by mainshocks (17) and aftershocks (19).

Then we defined **20 clusters** for the dendrogram analysis (Figure 4). We are able to well separate seismic events from ambient noise.

**Figure 3.** 20 dimensions feature space in function of time

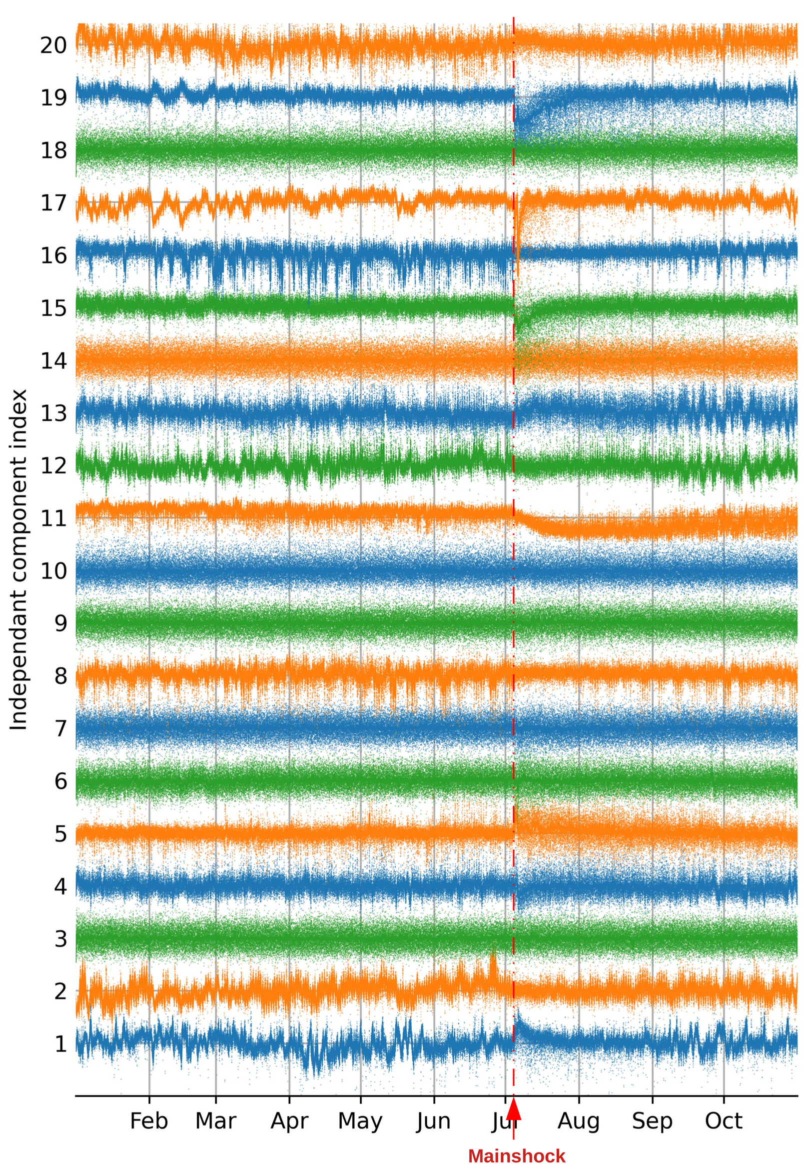
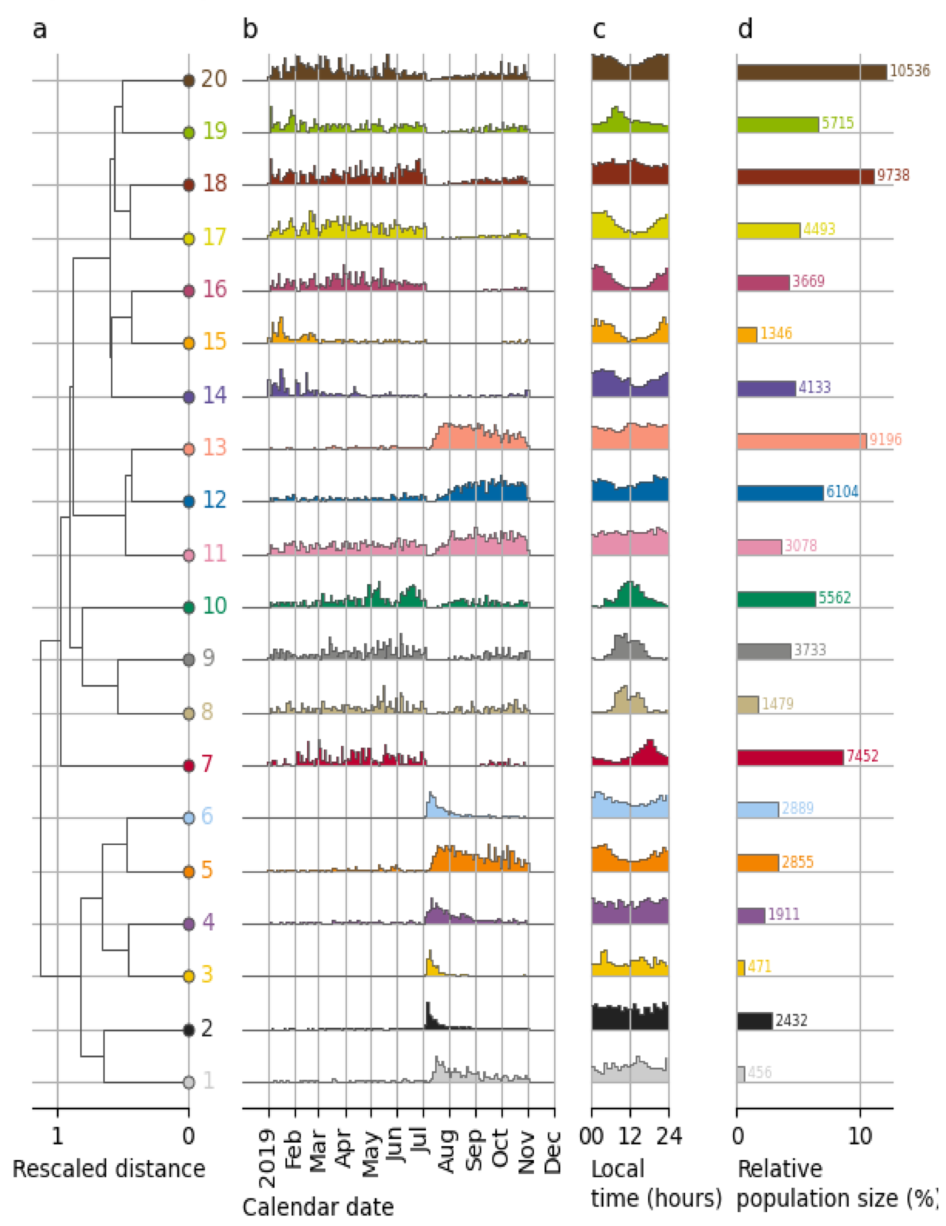



**Figure 4.** Dendrogram




### Cumulative number and localisation of events using SCSN seismic catalog from South-California

**Figure 5.** Seismic excess around Ridgecrest from 2016to 2019.

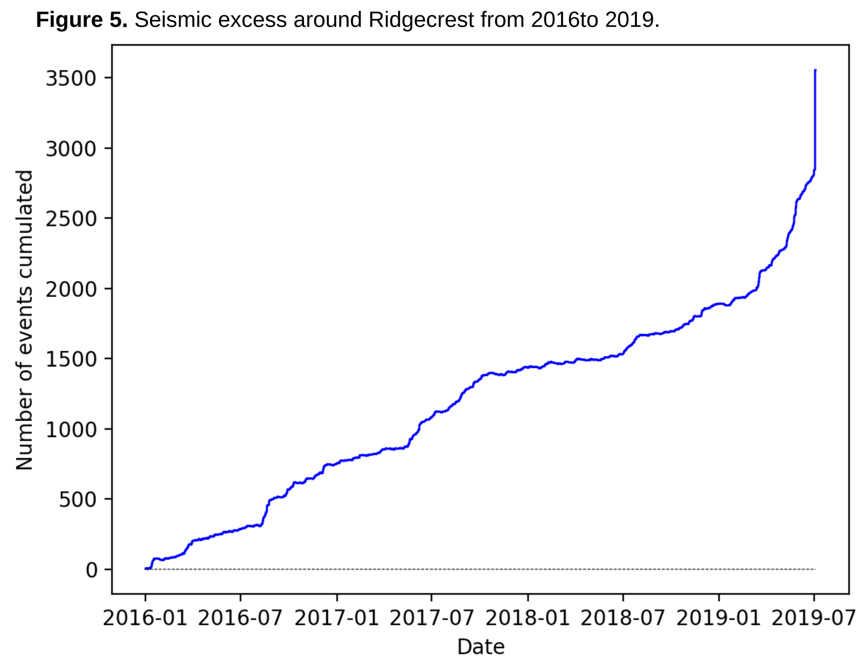



**Figure 6.** Seismic excess around Ridgecrest from 2016to 2019, for several distance to the hypocenter.

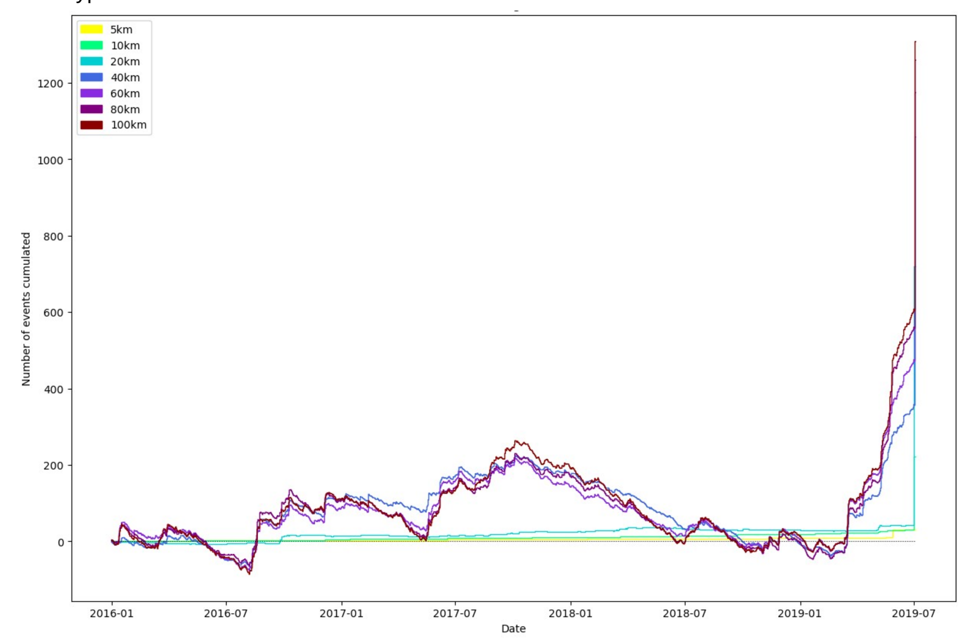
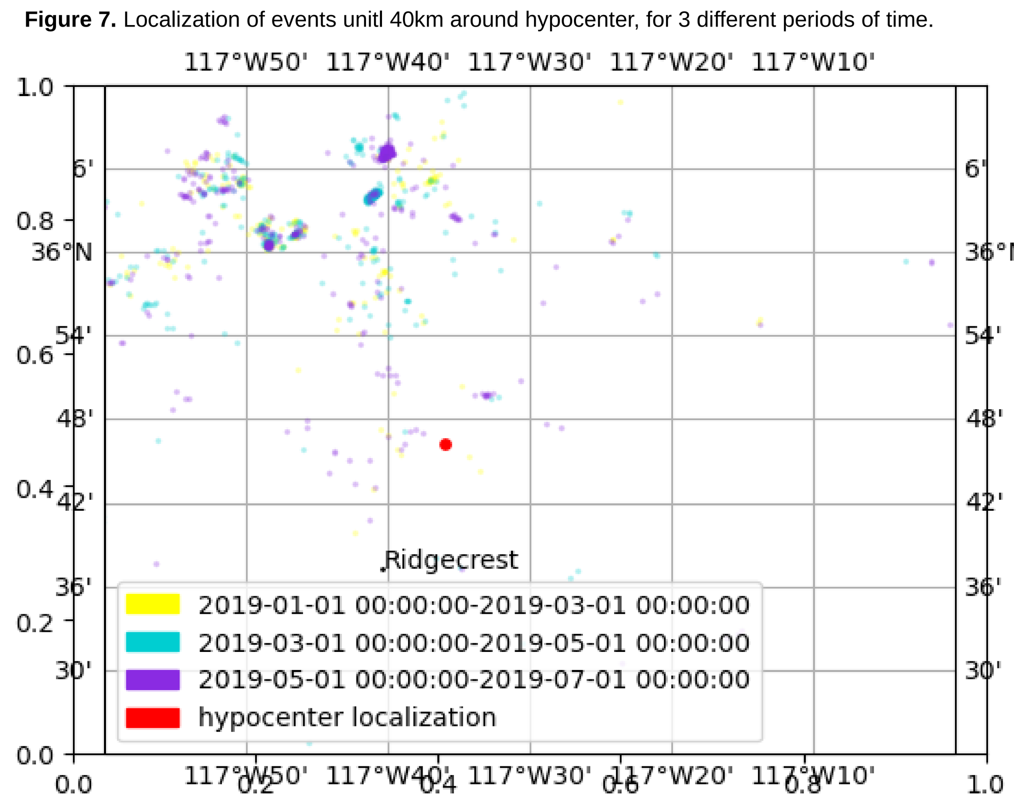


**Figure 7.** Localization of events until 40km around hypocenter, for 3 different periods of time.




We used the seismic catalog from SCSN, South-California to study seismicity around ridgecrest.

The cumulative number of events shows an important seismic excess few months before the 2019 earthquake (Figure 5). The figure 6 depicted that this excess is beyond at least 40km of the hypocenter.

The location of events before earthquake shows the existance of clusters of event North-West of the faults of the 2019 earthquake (Figure 7).

According to this results it seems that may exist precursors few kilometers North-West of the earthquake hypocenter. But it need to be proved.

## Outlook

Are we able to automate the search for early warning signs?

Using 10 years of data from B918 station closed to the previous clusters of events we want to know if we are able to detect the early seismicity using our clustering method as for the seismicity of the 2019 earthquake.

Moreover we have to take into account the difference with the hudge amount of aftershocks after the earthquake which can be easily detect and seperated using the culstering method and the few thousands of events detected in the catalog compared to all the noise's windows.