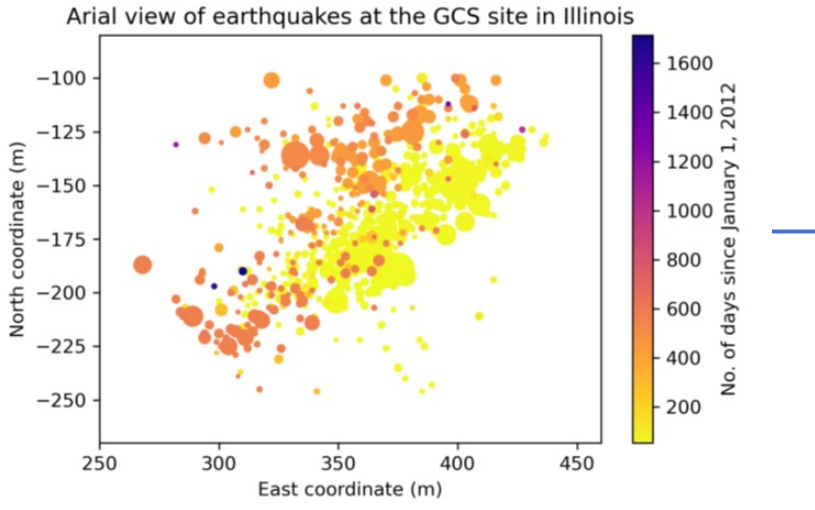
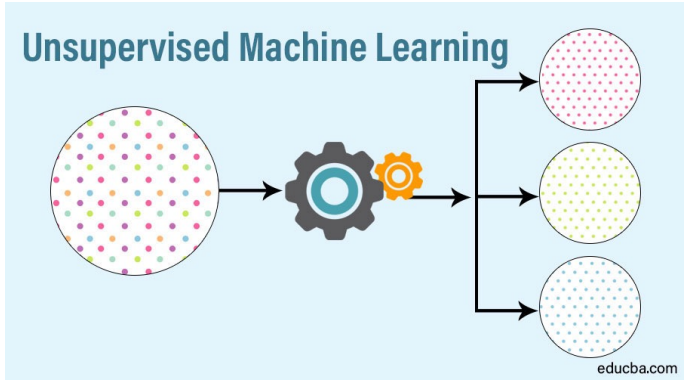
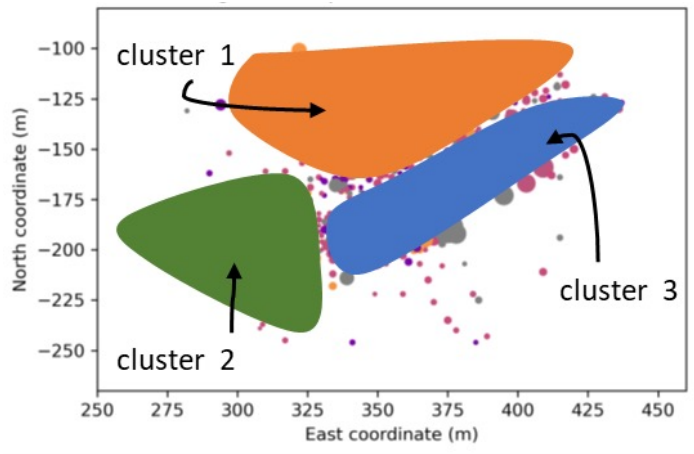
Clustering of carbon dioxide injection-induced earthquakes using feature-based representation of time series data

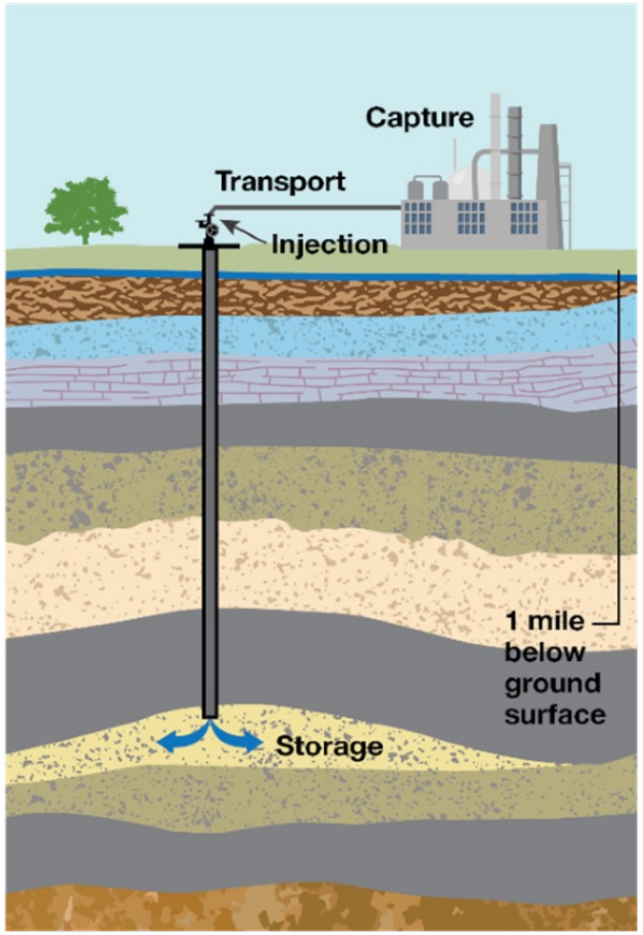
Runako Gentles, Lluís Saló-Salgado, Hannah Lu, David Castiñeira,
Michael Fehler and Ruben Juanes







Geological Carbon Sequestration (GCS) can help us achieve global net zero goals

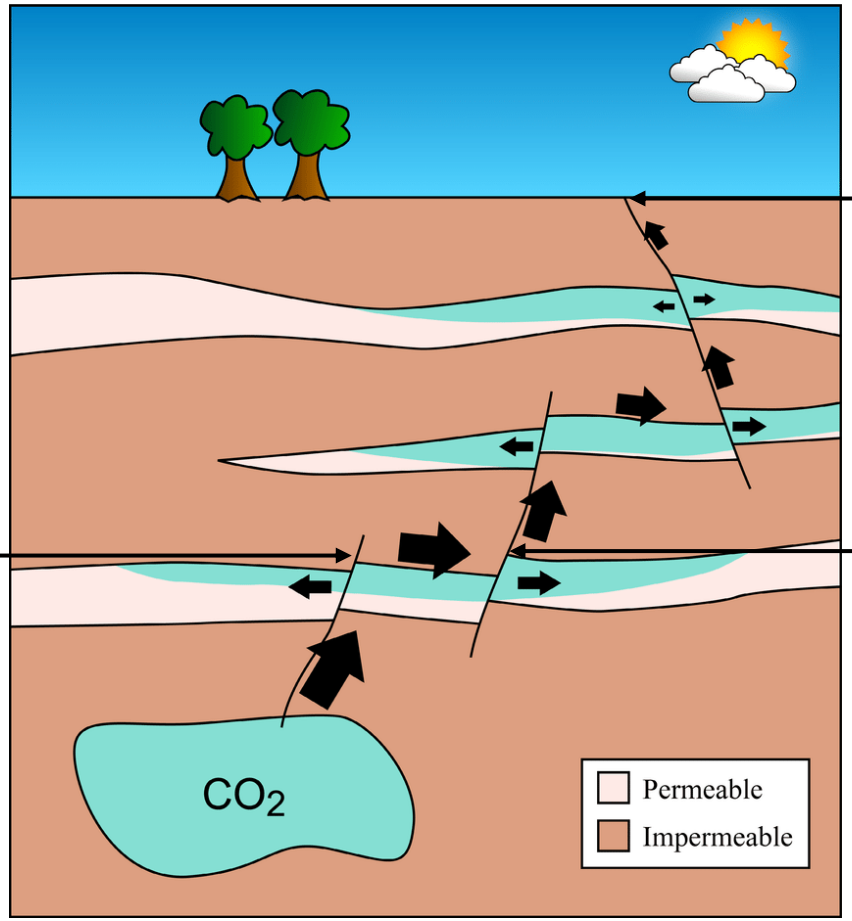


Class VI - Wells used for Geologic Sequestration of Carbon Dioxide, US EPA

GCS involves:

- Compressing captured CO_2 into a supercritical fluid (a dense but gas-like fluid)
- Transporting supercritical CO_2 to injection site
- Pumping supercritical CO_2 ~1000 m below the surface down a well and into a natural underground fluid reservoir (e.g. a saline aquifer)

Induced seismicity and leakage of CO_2 from storage reservoirs are among the risks associated with GCS



CO_2 escaping at high concentrations has health effects

Fault (fracture within the subsurface rock)

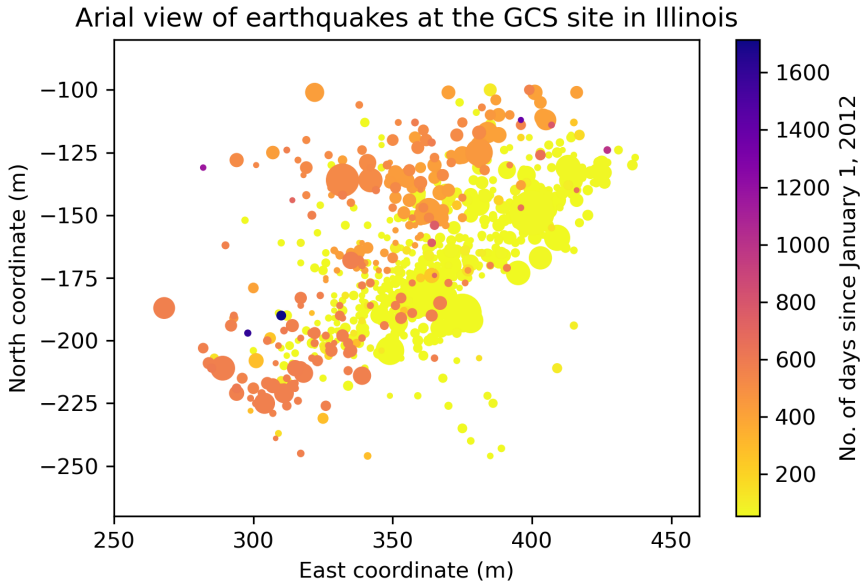
Escaped CO_2 can increase fault stress above the threshold needed for an earthquake

Won Chang, K. et al. *Modeling Leakage through Faults of CO_2 Stored in an Aquifer*

My aim is to use machine learning to analyze seismic time series to gain more insights into GCS seismic hazards

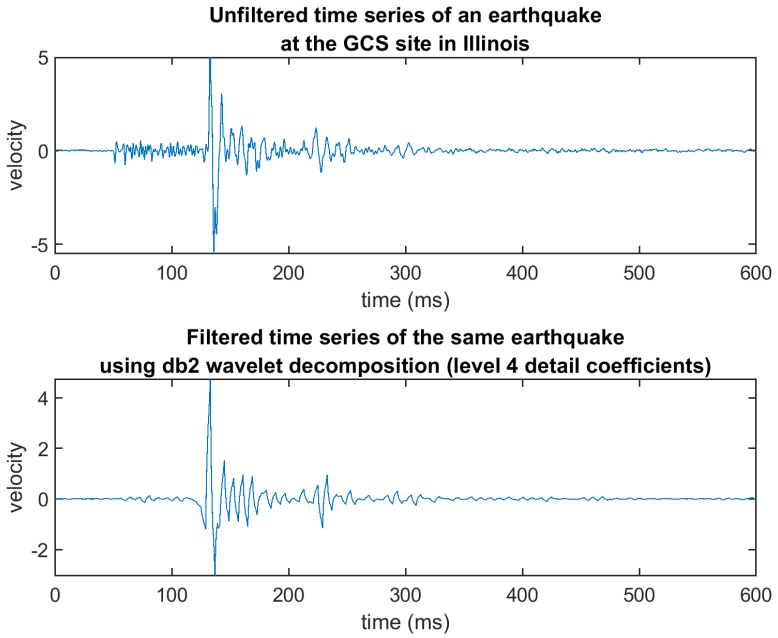
1

Preliminary analysis of 858 earthquakes from the GCS site in the Illinois Basin – Decatur Project



2

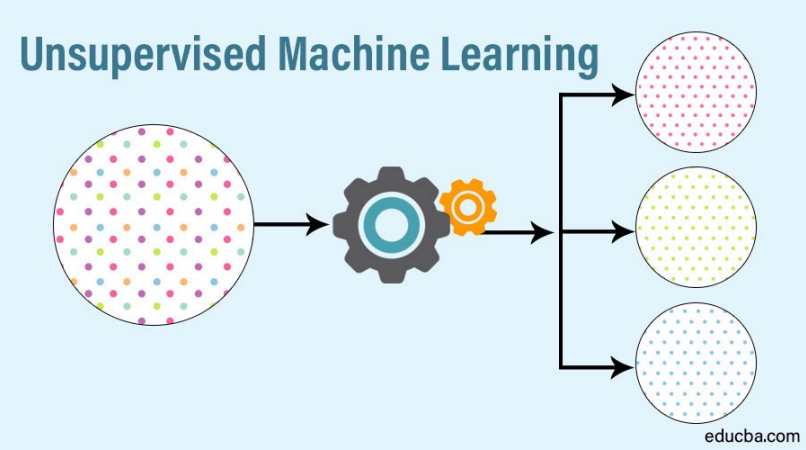
Filtering each time series with wavelet decomposition



My aim is to use machine learning to analyze seismic time series to gain more insights into GCS seismic hazards

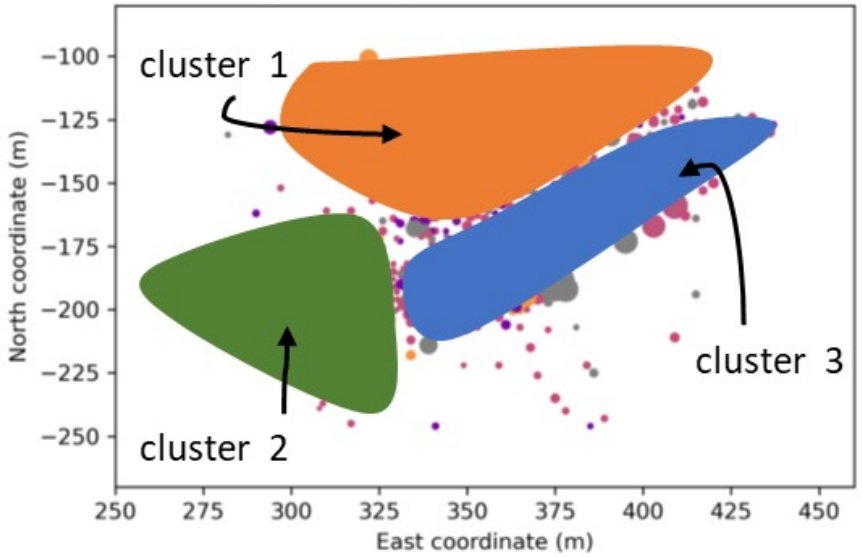
3

Reducing dimensionality of earthquake representation; clustering of this lower dimensional dataset



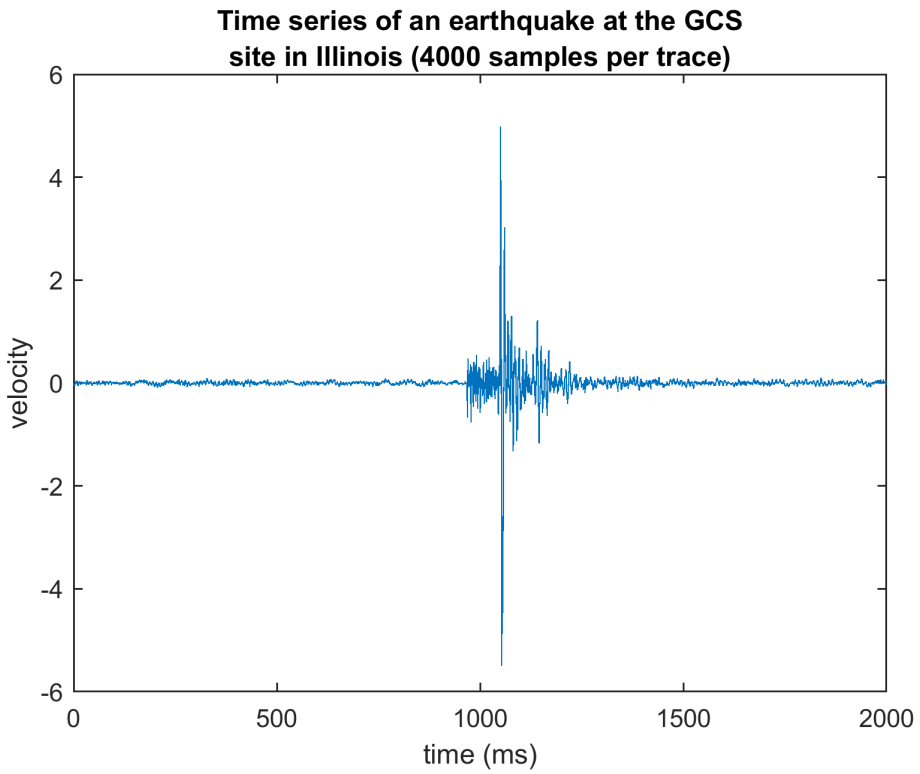
4

Identifying clusters



We significantly reduced the dimensionality of our 858-earthquake dataset

4000-dimensional time series

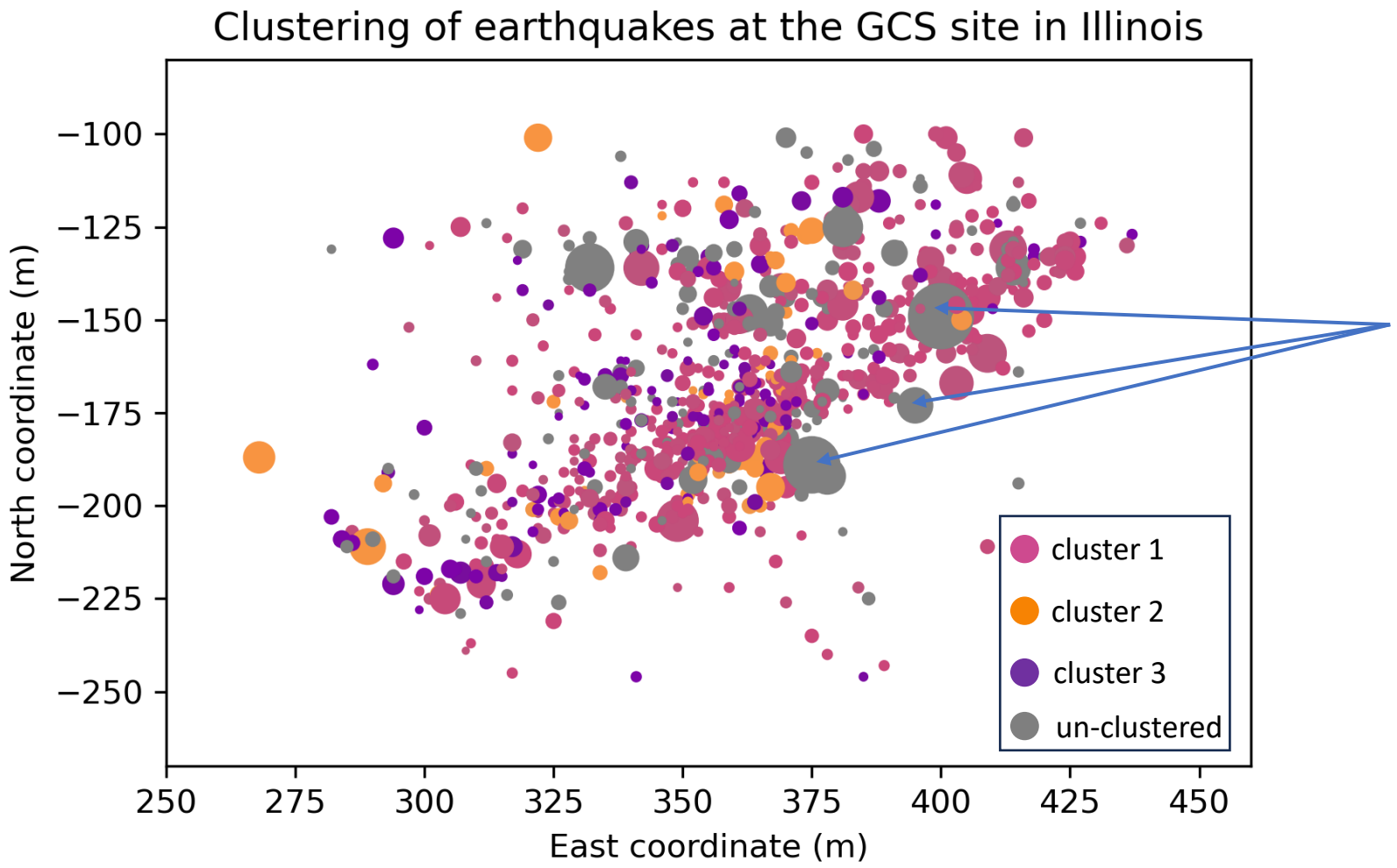


120-dimensional set of exemplary statistical values

120x1 cell

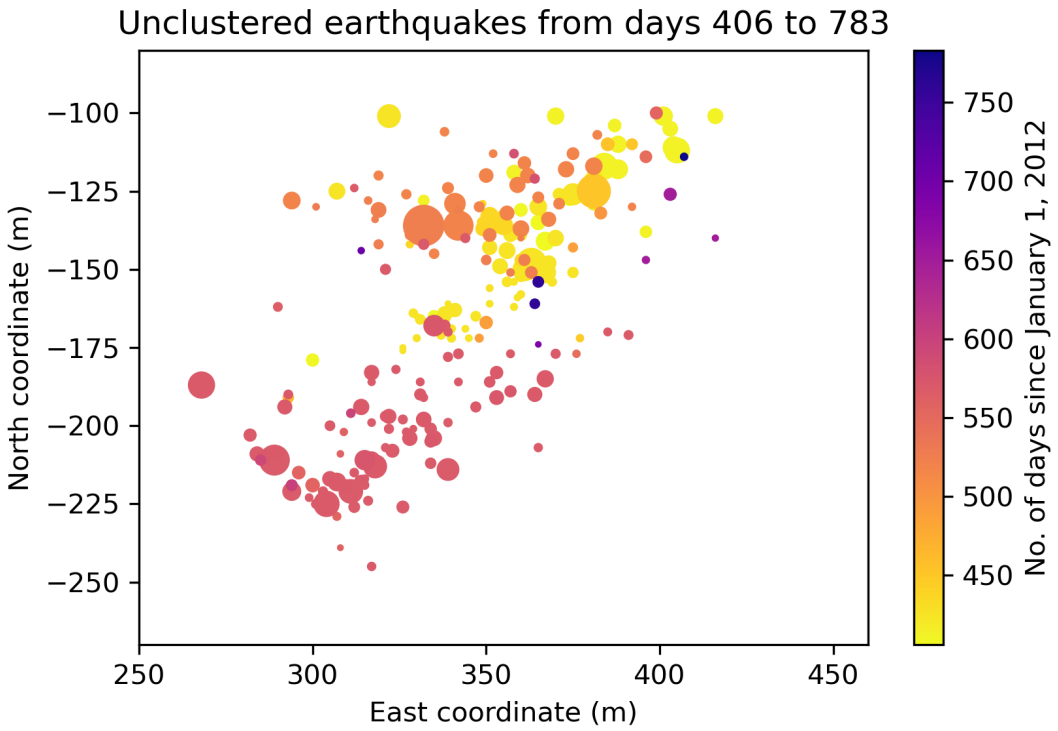
	1
1	DN_Mean(x,'norm')
2	DN_Withinp(y,0.5)
3	DN_Quantile(y,0.96)
4	ST_SimpleStats(y,'zcross')
5	DN_ProportionValues(x,'geq0')
6	CO_AutoCorr(y,2)
7	CO_AutoCorr(y,11)
8	CO_AutoCorr(y,37)
9	SY_StatAv(y,'seg',6)
10	SY_SlidingWindow(y,'std','apen',5,2)
11	SY_SlidingWindow(y,'ent','ent',2,1)
12	SY_SlidingWindow(y,'apen','ent',2,10)
13	SY_SlidingWindow(y,'apen','ent',5,2)
14	SY_SlidingWindow(y,'apen','ent',10,2)
15	SY_SlidingWindow(y,'lillie','apen',5,1)
16	SY_SlidingWindow(y,'lillie','apen',10,10)
17	SY_SlidingWindow(y,'lillie','ent',10,2)
18	SY_SlidingWindow(y,'AC1','ent',2,1)
19	SY_LocalGlobal(y,'l',10)
20	SY_LocalGlobal(y,'l',100)

Clustering lower-dimensional earthquake set identified large magnitude events

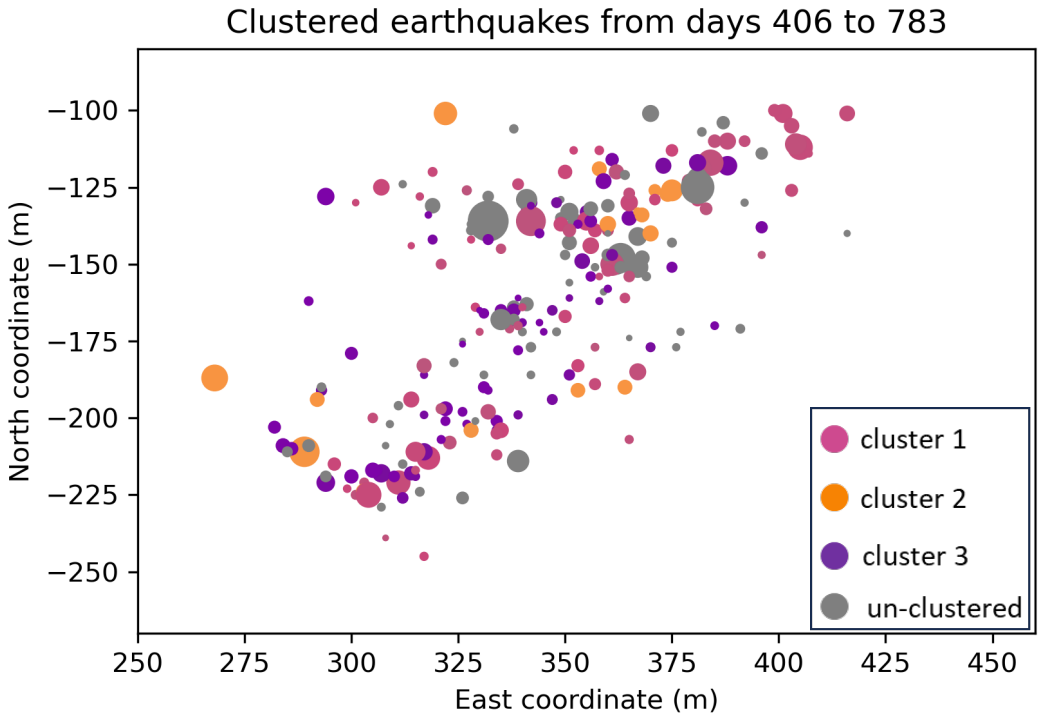


Several larger magnitude earthquakes remain un-clustered (represented in grey)

We are tuning our clustering to identify the obvious time distribution of the earthquakes



Section of earthquake plot
colored based on day of occurrence



Section of earthquake plot
colored based on clustering result

Next steps

- Refine the clustering process to better observe the time distribution and separation of large magnitude earthquakes from smaller ones

- Write a report on the effectiveness of representing these earthquakes in a lower dimension with our methodology