Physics-Guided Data Science for Earthquake Detection

Lei Huang, Nishath Ranasinghe, Ted Clee

Computer Science Department
Prairie View A&M University
lhuang@pvamu.edu

Fourth MCBDA Workshop
May 14, 2019
Prairie View A&M University

CREDIT Center Research Thrust Team 1

- **Faculty**
  - Dr. Lei Huang
  - Dr. Barbara Chapman
- **Graduate Research Assistants**
  - Myron Garner
  - Isaiah Brown
  - Hao Zheng
  - Sokunbi Oluwasomidotun
  - Miguel Polanco
- **Acknowledgements**
  - Dr. Nishath Rajiv Ranasinghe, Postdoctoral Fellow
  - Dr. Ted Clee, Research Scientist
  - Jeremy Kemp, Research Associate

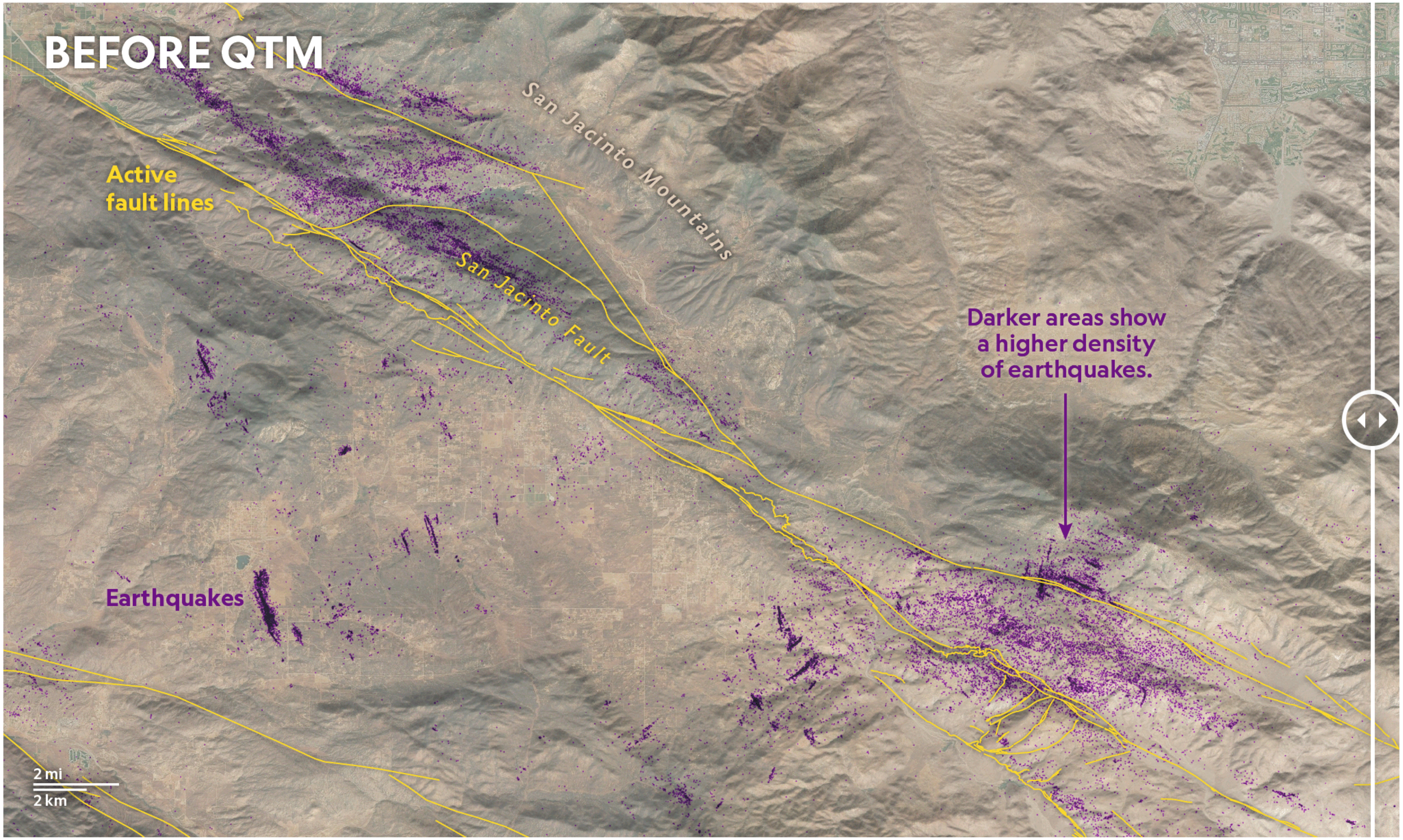
Why study earthquakes?

- **Study Earth structure** – seismic wave propagation reveals global structure and local crustal variations
- **Public safety** – early warning of quakes and associated tsunamis can mitigate disaster effects
- **Nuclear Surveillance** – distinguish enemy underground nuclear tests
- **Hazard monitoring** – man-made explosions and induced seismicity from water injection may cause local hazards

Seismic event types in Earth’s crust

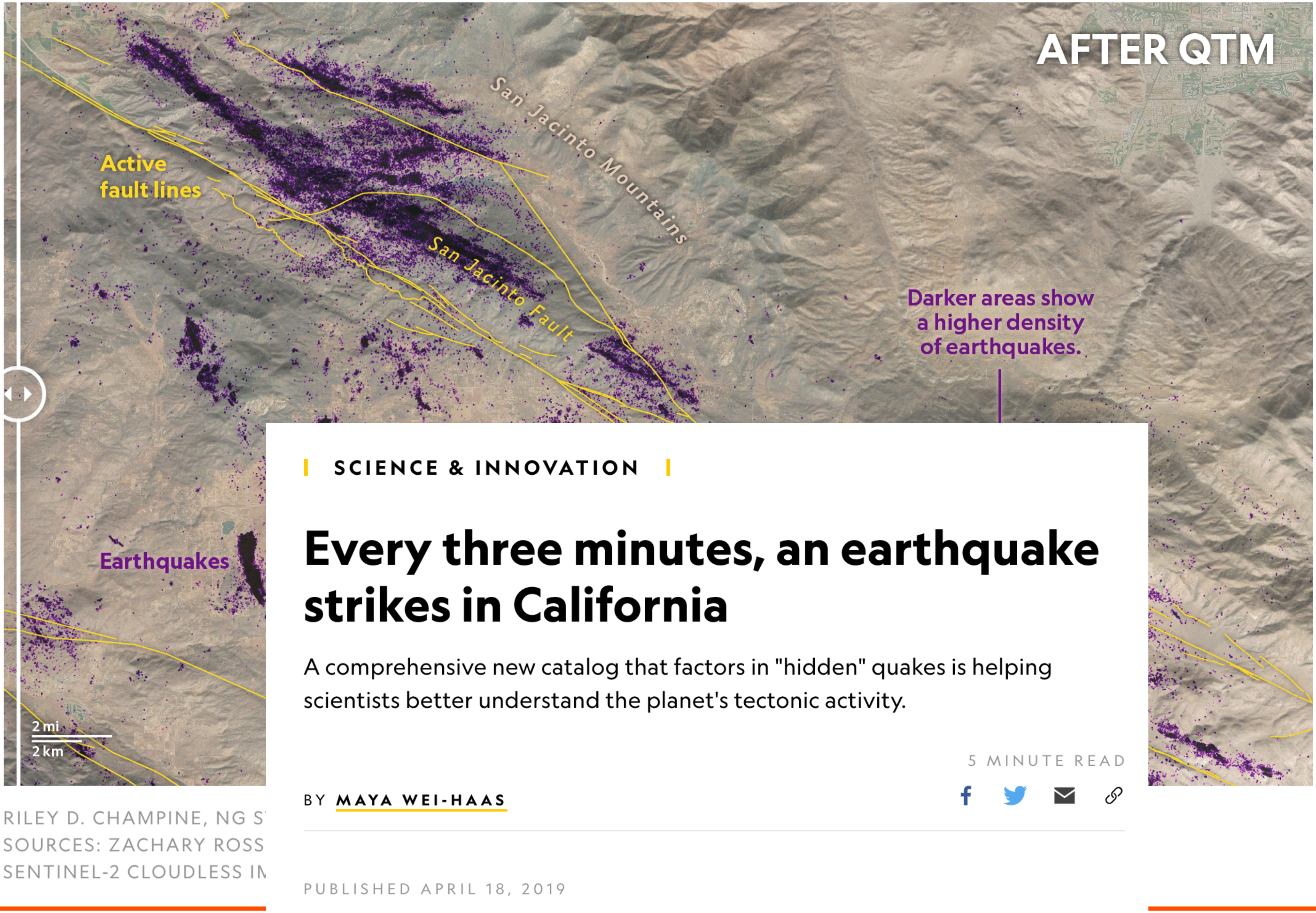
- **Natural earthquakes** – fault slippage at depth, in areas of known tectonic activity, unknown timing
- **Nuclear test explosions** – spherically symmetric impulse at drillable depth, in geopolitically sensitive locations, secret timing
- **Borehole shots** – Localized, controlled-source experiments for hydrocarbon exploration or geo-technical engineering purposes, times and locations known
- **Blasts from mining operations** – hemispherical radiation patterns at Earth surface, times and locations not recorded
- **Microseismics** – seismic energy generated from hydraulic fracturing along horizontal boreholes
- **Induced seismicity** from waste-water injection – similar to small earthquakes, in regions of drilling activity
- **Surface noise** sources – any motion on land surface, from truck vibrations to meteor impacts

Data analytics challenge – Earthquake discrimination



RILEY D. CHAMPINE, NG STAFF
SOURCES: ZACHARY ROSS, CALTECH; USGS; GLOBAL EARTHQUAKE MODEL;
SENTINEL-2 CLOUDLESS IMAGERY BY EOX IT SERVICES GMBH

Data analytics challenge – Earthquake discrimination



RILEY D. CHAMPINE, NG S
SOURCES: ZACHARY ROSS
SENTINEL-2 CLOUDLESS IN

SCIENCE & INNOVATION

Every three minutes, an earthquake strikes in California

A comprehensive new catalog that factors in "hidden" quakes is helping scientists better understand the planet's tectonic activity.

5 MINUTE READ

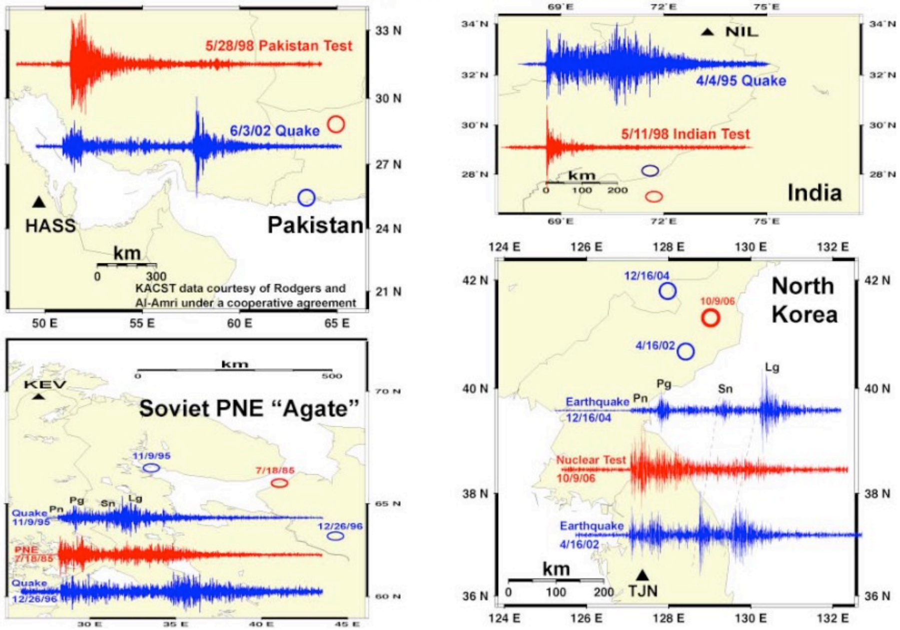
BY **MAYA WEI-HAAS**

PUBLISHED APRIL 18, 2019

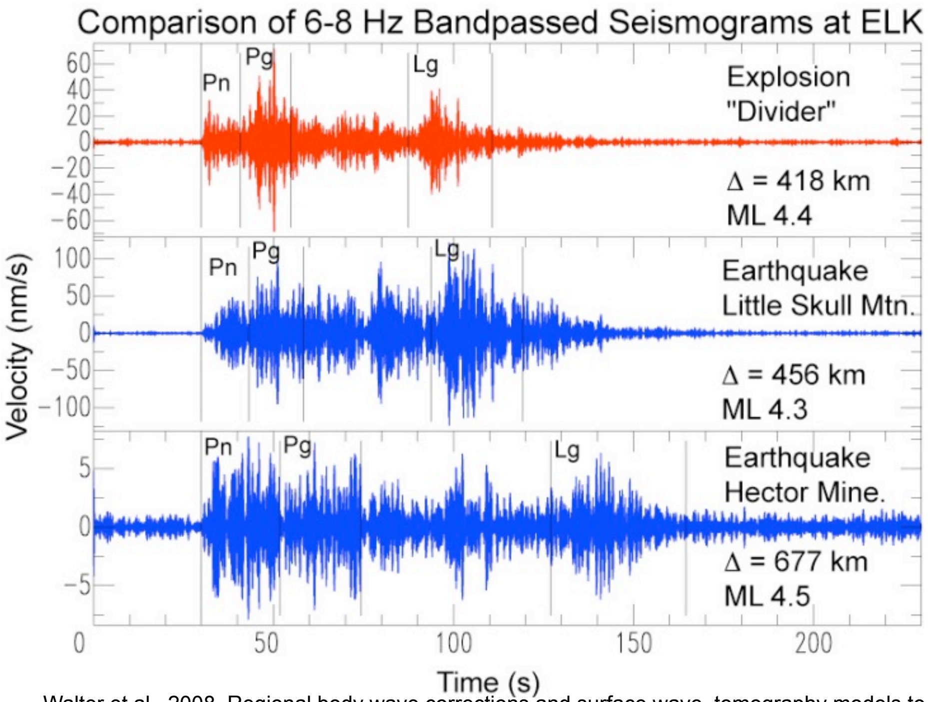
Data analytics challenge – Nuclear discrimination

High-frequency P/S appears to work everywhere

6-8 Hz **Earthquake** and **Explosion** pairs



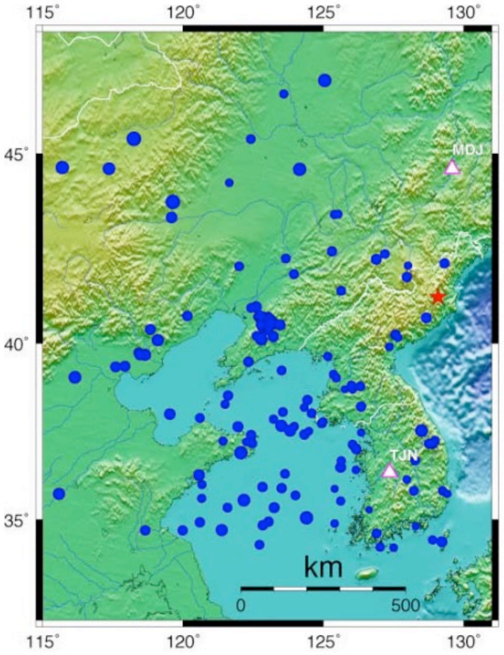
Walter et al. 2008, Regional Seismic Amplitude Modeling and Tomography for Earthquake-Explosion Discrimination

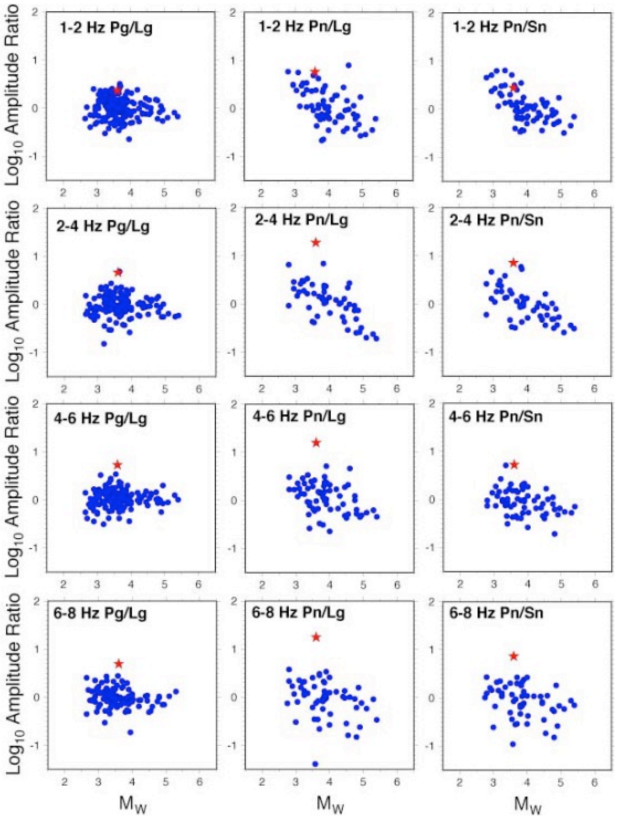


Walter et al., 2008. Regional body wave corrections and surface wave tomography models to improve discrimination.

- Explosions are difficult to discriminate from earthquakes in real time for monitoring purposes.
- In general, explosions produce proportionally more P wave energy than earthquakes.

Differentiating explosions from earthquakes using amplitude ratios



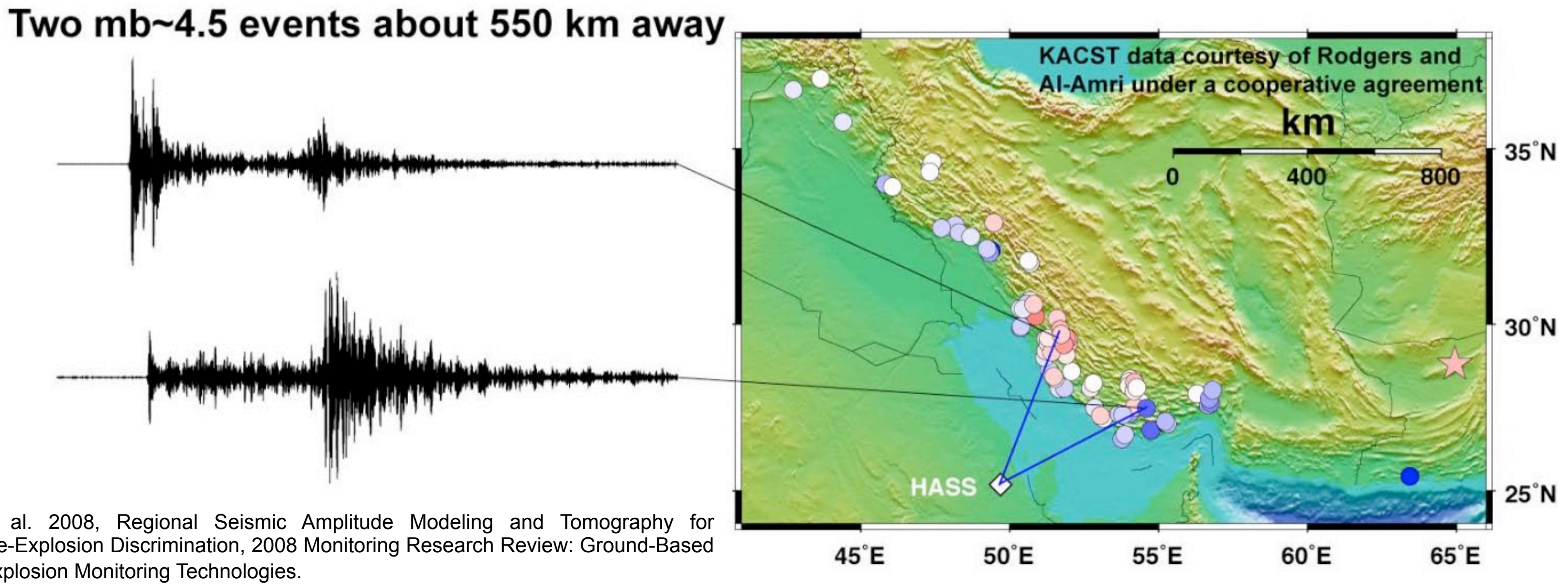


Walter et al. 2008, Regional Seismic Amplitude Modeling and Tomography for Earthquake-Explosion Discrimination, 2008 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies.

- Explosions may be identified by comparing wave forms originating from an explosion against previously identified earthquakes from the same region at different frequency bands.

Data analytics challenge – Nuclear discrimination

Example of large 6-8 Hz Pn/Sn variation in the Middle East



Walter et al. 2008, Regional Seismic Amplitude Modeling and Tomography for Earthquake-Explosion Discrimination, 2008 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies.

- Use **more seismic data** collected from **different regions** of the globe to train the model and improve the predictability.
- Use latest techniques to improve real time **event detection** and **physical property** calculation.
- Train the model using data collected from **nuclear explosions**.

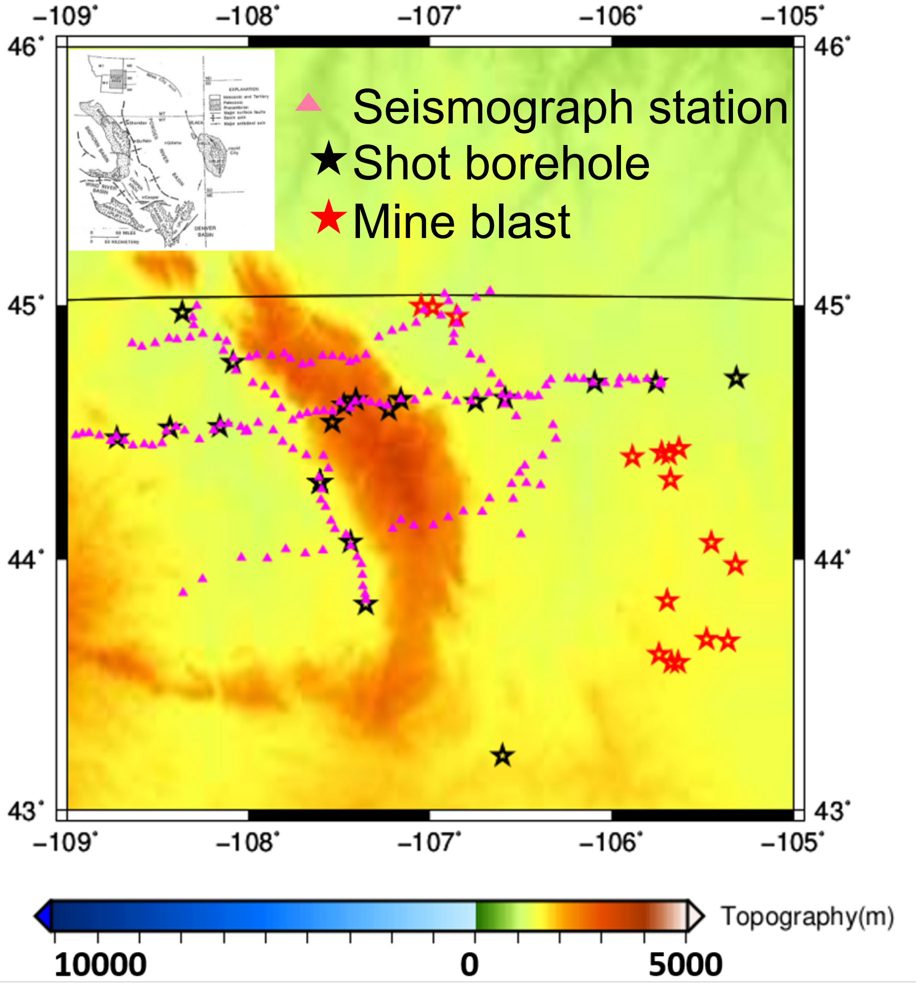
Detecting and analyzing seismic events using physical characteristics

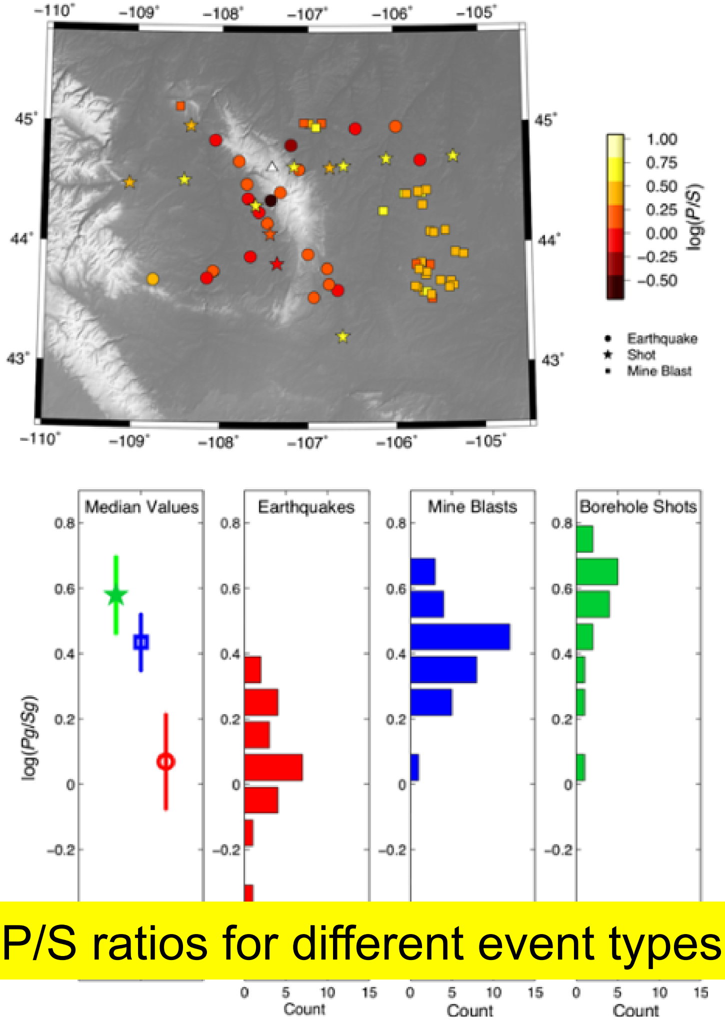
- **Location – hypocenter on Earth surface**
  - Precisely determined by triangulation of first arrivals at 3 or more stations
  - For a single station, approximate azimuth from ratio of two horizontal components, and distance from time separation of fast P and slower S arrivals
- **Depth – epicenter at depth below surface**
  - Obtained by careful analysis of additional phases, such as pP arrival reflected once from surface
- **Size**
  - Weaker events need more sophisticated effort to detect above noise
- **Focal mechanism**
  - Fault slip mode is distinguished from explosion by different waveform characteristics
  - Use “fingerprint” method to compare to known events

Machine learning case study for event classification

- We seek to train a neural network to **classify** events as **earthquakes, borehole shots, mine blasts,** or **noise**
- Data set for the case study consists of multiple seismograph **station recordings**, with **cataloged events** labeled for each source type
- Study area is in the Powder River basin of North-east Wyoming, an area of active **coal mining** and **petroleum exploration**
- The **172** short-period seismograph stations in the Bighorn seismic array also provided **earthquake** event and **noise** data from 5 weeks of earthquake monitoring
- A possible **event type discriminator** is the **P/S ratio** of P-wave and S-wave arrival amplitudes

Bighorn Seismic Array



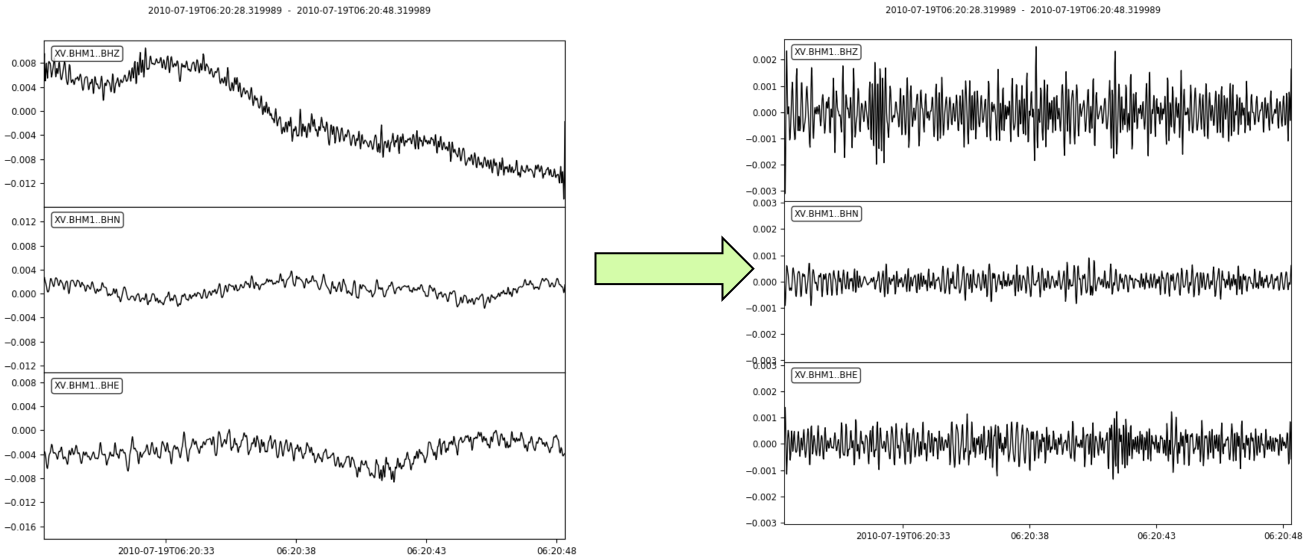


P/S ratios for different event types

O’Rourke et al. 2016, *BSSA*,106 (5),2320–2331

Data Pre-processing

- **Extract events from raw seismic based on the catalog**
  - Create a 10-second window for each event
  - Calculate the event arrival time based on the event distance and average velocity
- **Create noise data based on the catalog**
  - Avoid the event windows
- **Preprocess data**
  - Detrend
  - Normalize
  - Filter
  - Resample



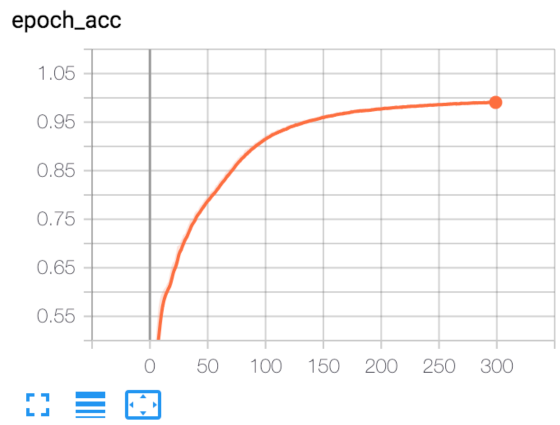
Deep learning network

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
===
input_1 (InputLayer)        (None, 1001, 3)          0
_________________________________________________________________
conv1d (Conv1D)              (None, 1001, 32)         320
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 1001, 32)         3104
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 500, 32)          0
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 500, 32)          3104
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 500, 32)          3104
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 250, 32)          0
_________________________________________________________________
conv1d_4 (Conv1D)            (None, 250, 32)          3104
_________________________________________________________________
conv1d_5 (Conv1D)            (None, 250, 32)          3104
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 125, 32)          0
_________________________________________________________________
conv1d_6 (Conv1D)            (None, 125, 32)          3104
_________________________________________________________________
conv1d_7 (Conv1D)            (None, 125, 32)          3104
_________________________________________________________________
max_pooling1d_3 (MaxPooling1 (None, 62, 32)           0
_________________________________________________________________
conv1d_8 (Conv1D)            (None, 62, 32)           3104
_________________________________________________________________
conv1d_9 (Conv1D)            (None, 62, 32)           3104
_________________________________________________________________
max_pooling1d_4 (MaxPooling1 (None, 31, 32)           0
_________________________________________________________________
```

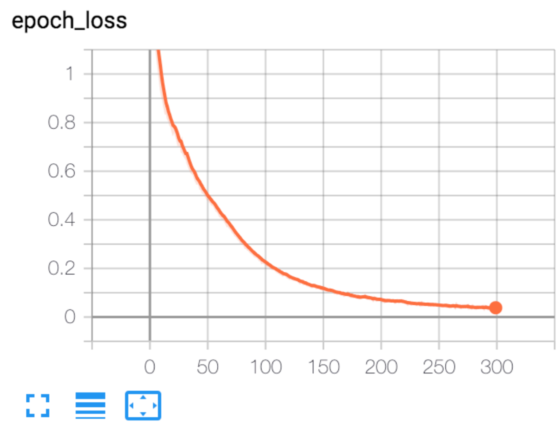
```
conv1d_10 (Conv1D)           (None, 31, 32)           3104
_________________________________________________________________
conv1d_11 (Conv1D)           (None, 31, 32)           3104
_________________________________________________________________
max_pooling1d_5 (MaxPooling1 (None, 15, 32)           0
_________________________________________________________________
conv1d_12 (Conv1D)           (None, 15, 32)           3104
_________________________________________________________________
conv1d_13 (Conv1D)           (None, 15, 32)           3104
_________________________________________________________________
max_pooling1d_6 (MaxPooling1 (None, 7, 32)            0
_________________________________________________________________
conv1d_14 (Conv1D)           (None, 7, 32)            3104
_________________________________________________________________
conv1d_15 (Conv1D)           (None, 7, 32)            3104
_________________________________________________________________
max_pooling1d_7 (MaxPooling1 (None, 3, 32)            0
_________________________________________________________________
dropout (Dropout)            (None, 3, 32)            0
_________________________________________________________________
flatten (Flatten)            (None, 96)               0
_________________________________________________________________
dense (Dense)                (None, 512)              49664
_________________________________________________________________
dropout_1 (Dropout)          (None, 512)              0
_________________________________________________________________
dense_1 (Dense)              (None, 4)                2052
=================================================================
====
Total params: 98,596
Trainable params: 98,596
Non-trainable params: 0
_________________________________________________________________
```

Results

epoch_acc



epoch_loss



Prediction for new dataset:

Loss	**Accuracy**
0.425	0.746719

Classes	**ID**
Noise	0
Earthquake	1
Active Shot	2
Mine Blast	3

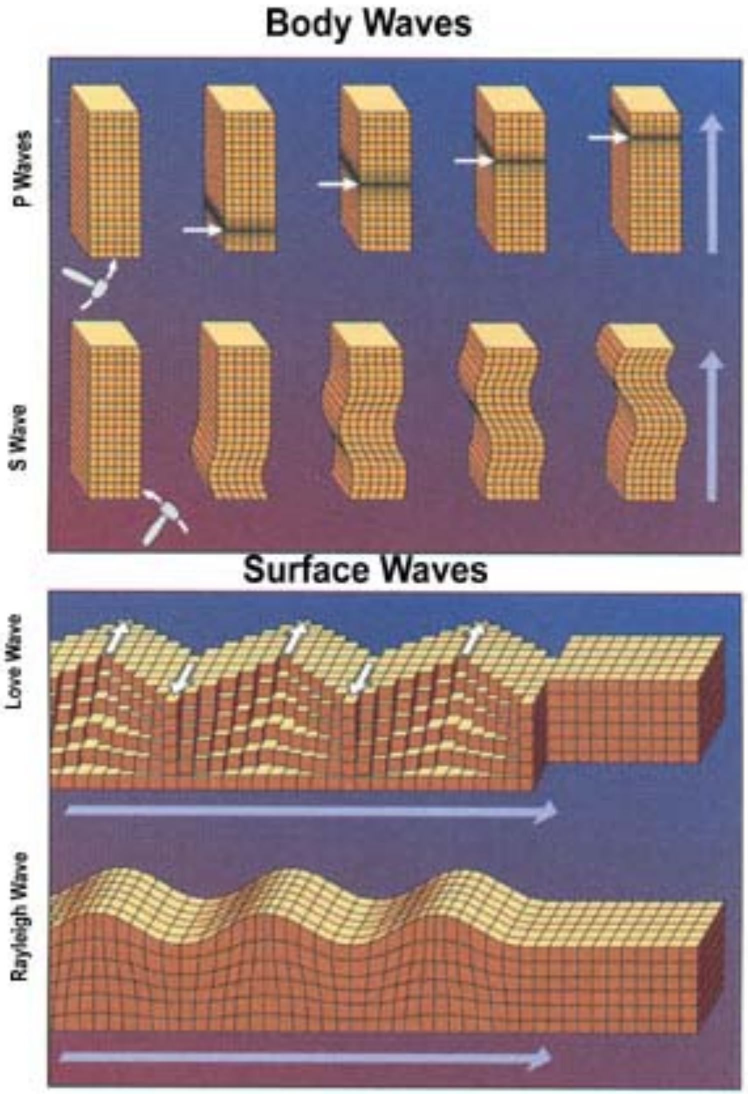
Takeaway Points

- It is challenging to detect small/subtle/micro seismic activities
- Data preprocessing is critical to achieve high-quality data analytics results
- Need a more accurate event extraction algorithm to detect the first arrivals of P and S waves
- Need to have better noise data that do not have any small (undetected) events included
- It is critical to follow the physics principles in data analytics

Questions?

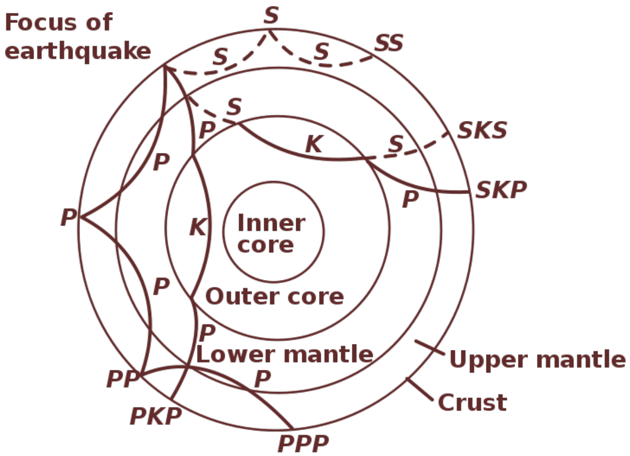
Elastic wave propagation modes

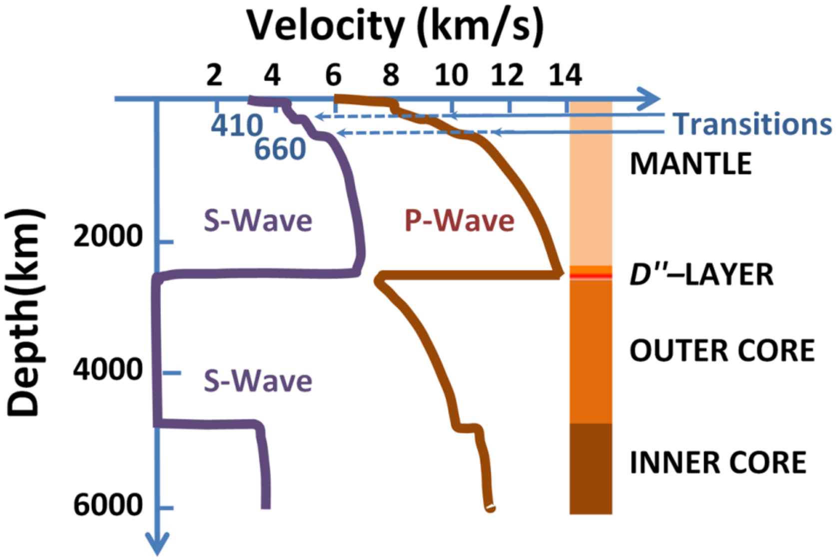
- **Body waves within a medium**
  - Compressional (P) waves travel as a “train” of compressions and rarefactions
  - Shear (S) waves travel with side-to-side motion
- **Surface waves**
  - Love waves move horizontally, transverse to propagation direction
  - Rayleigh waves move vertically, traveling along the surface

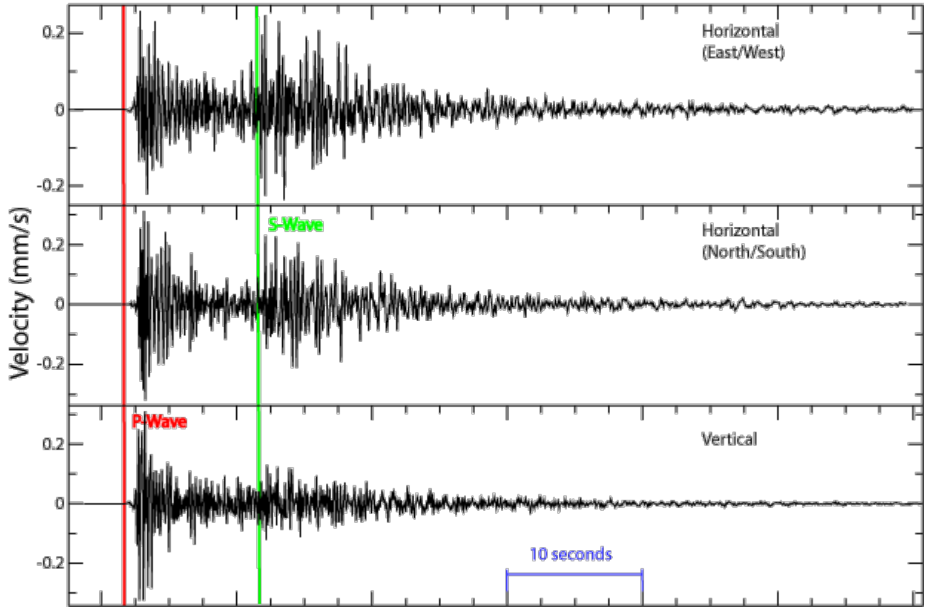


(USGS)

Earth structure deduced from arrivals of seismic phases







Vertical and two horizontal components are recorded at each seismograph station. Note that up-going S arrival is weaker on vertical channel. Complexity of arrivals comes from details of source activation and propagation path.