Regional Seismic Discrimination using Machine Learning

AKSHAY ARAVAMUDAN[1], HAFIDH A A GHALIB[2], GEORGIOS C. ANAGNOSTOPOULOS[1], ANTHONY O. SMITH[1], MATTHEW SIBOL[2], GORDON KRAFT[2], WILMER RIVERS[2], YAJUN WANG[2], XI ZHANG[1], RUKSANA KABEALO[1] AND ERBENE DE CASTRO MAIA JUNIOR[1]

(1) FLORIDA INSTITUTE OF TECHNOLOGY, ELECTRICAL ENGINEERING AND COMPUTER SCIENCE, MELBOURNE, UNITED STATES,

(2) ARRAY INFORMATION TECHNOLOGY, MELBOURNE, UNITED STATES

THIS MATERIAL IS BASED UPON WORK SUPPORTED BY DEFENSE THREAT RESEARCH ORGANIZATION (DTRA) CONTRACT NO. HDTRA1-22-C-0005, “A DEEP LEARNING APPROACH FOR ENHANCED IDENTIFICATION OF NUCLEAR EXPLOSIONS (DELAEINE),”. ANY OPINIONS, FINDINGS, CONCLUSIONS, OR RECOMMENDATIONS CONTAINED HEREIN ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REPRESENT THE OFFICIAL POLICIES OR ENDORSEMENTS, EITHER EXPRESSED OR IMPLIED, OF DTRA OR THE U.S. GOVERNMENT.

Introduction

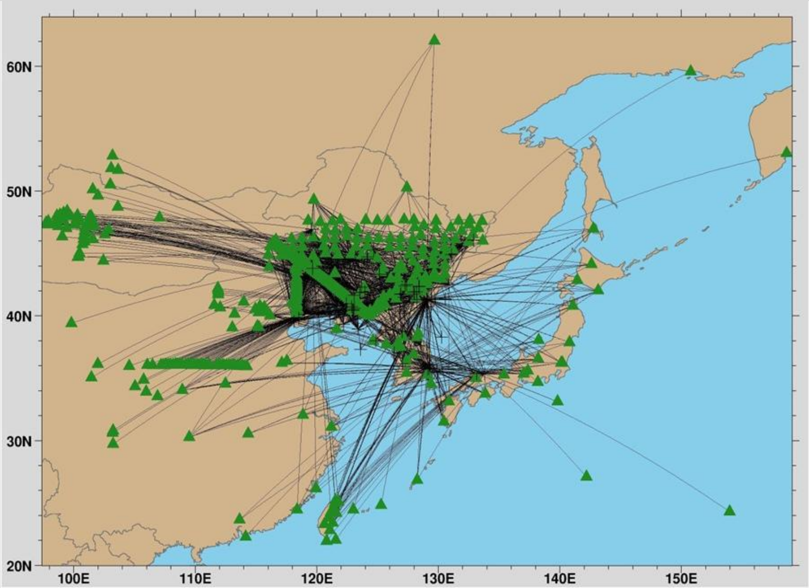
- **Motivation**
  - Distinguishing between earthquakes and nuclear explosions at regional scales is a complex geophysical problem.
  - At epicentral distances (< 2,000 km), signals propagate through the crust and deteriorate in quality, which makes the discrimination task harder.
  - While there are sophisticated statistical discriminant-based methods used in practice,
    - They require human intervention to produce predictions from P/S ratios.
    - This has been effective for short range explosions mining, rockburst and open-pit explosions, but it becomes harder to produce discriminants at regional scales due to the inconsistent medium of travel.

- **To remedy this**, we use machine-learning models whose data-driven approach allows to learn complicated decision boundaries between the two events types.

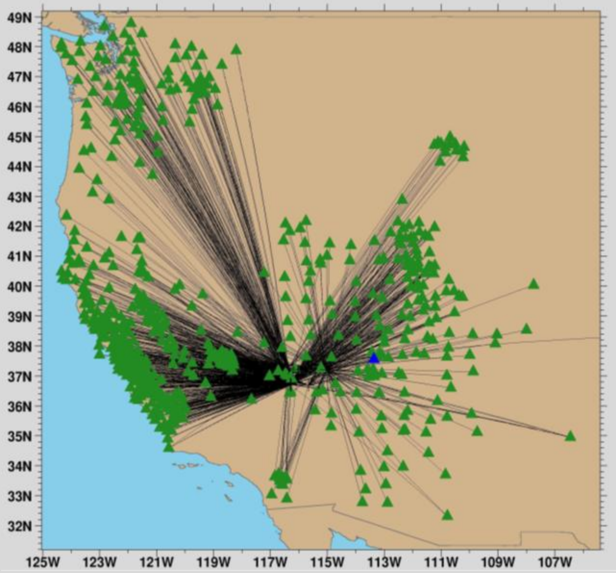
- **In this work**, for events recorded at regional distances located in western USA and eastern Asia, we
  - Investigate the use of various ML models and feature transformation methods in discriminating between earthquakes and nuclear explosions.
  - Study the effect of multiple channels on discrimination performance.
  - Propose and evaluate a novel geophysics inspired feature transformation – slonograms – to adjust for epicentral distances.
  - Formulate a network discrimination framework to find consensus among individual station predictions.

Seismic Data

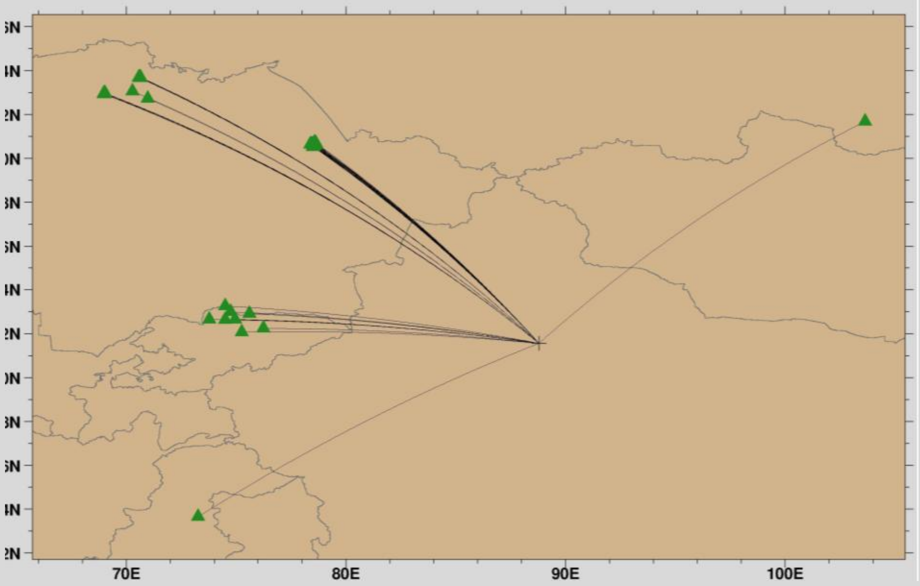
❑ North Korea



❑ Nevada



❑ Lop Nor



Seismic Data

Site/Source	No. of Events	No. of Explosions	No. of Earthquakes	No. of Unknown Events	No. of Waveforms Utilized	No. of Waveforms Rejected
North Korea	74	12	46	16	5502	148445
Nevada	50	37	12	1	7293	30291
Lop Nur	24	10	14	0	3111	23383

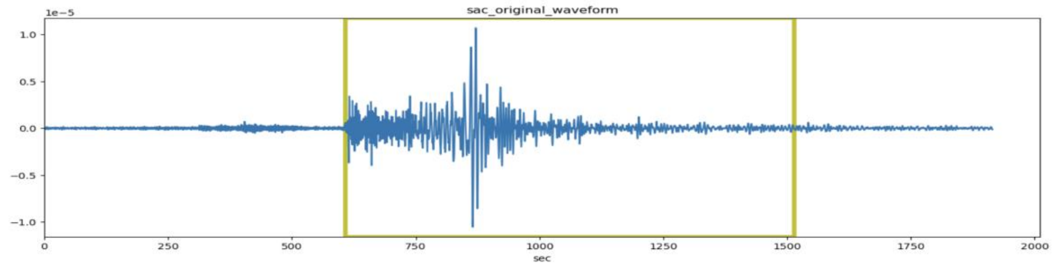
Regional (epicentral distance < 2000 km) waveform data analyzed

- **Waveform data source**
  - IRIS, SMU (Southern Methodist University), Waveforms From Nuclear Explosions (WFNE) and International Monitoring System (IMS).
  - All waveform data (SAC format) were manually analyzed, and the phase picks were reviewed.
- **Routine data preprocessing**
  - Removed the mean, trend and instrument response using poles-zero or response tables, as available.
  - P-wave start time was manually annotated.
  - Transformed the Z, NS, EW components to Z, R (Radial) and T (Transverse) components (channels).
  - Applied different filters when examining the waveforms. (high pass at 0.5Hz & low pass – near Nyquist frequency)

Data Quality Control (QC) Approach

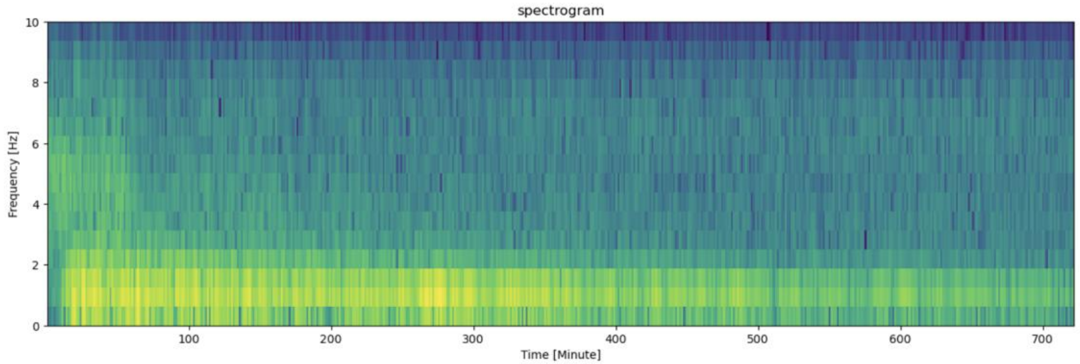
- Individual waveforms **were examined**, **characterized and cleared for ML processing.**
- **The SAC file header field KUSER0 used to classify the quality of the signal.**
- **For example, consider data with one or more spectral tones in their spectrograms.**
  - When one or more spectral tones are detected in a signal, the SAC file header field KUSER0 is set to “TONE”.
  - “TONE” is a relatively constant frequency of about 1 Hz in width that can change frequency as the noise source turns on or off
  - “TONE” is a noise that should not be used to classify other spectral signals like doppler shift signals.
- **QC descriptor definitions for KUSER0 are:**
  - OTHER → data rejected due to clipping, dropouts, mixed signals, large glitches, very low P-wave energy SNR, etc.
  - GOOD → clear P-wave signals from Pf to event end-time with no temporal or spectral issues being observed
  - LOWSNR → low Pf SNR can lead to IASPEI based arrival time being used for Pf, very low P-wave energy can result in rejection of the signal (*i.e.*, KUSER0 = OTHER)
  - GLITCHES → electronic or extraneous signals in the data; large or excessive glitches will result in rejection of the signal (i.e., KUSER0 = OTHER)
  - TONE → narrow (*e.g.*, ~1 Hz) constant frequency noise in signal spectrogram observed.
- **Summary:**
  - Data QC characterized as OTHER did not delivered to FIT
  - Data QC characterized as GOOD, LOWSNR, GLITCHES, or TONE delivered to FIT
  - The QC label (KUSER0) hierarchy is LOWSNR (low), GLITCHES, and TONE (high). For example,
    - A signal with LOWSNR and GLITCHES will have KUSER0 = GLITCHES
    - A signal with GLITCHES and TONES will have KUSER0 = TONE

Feature Transformation



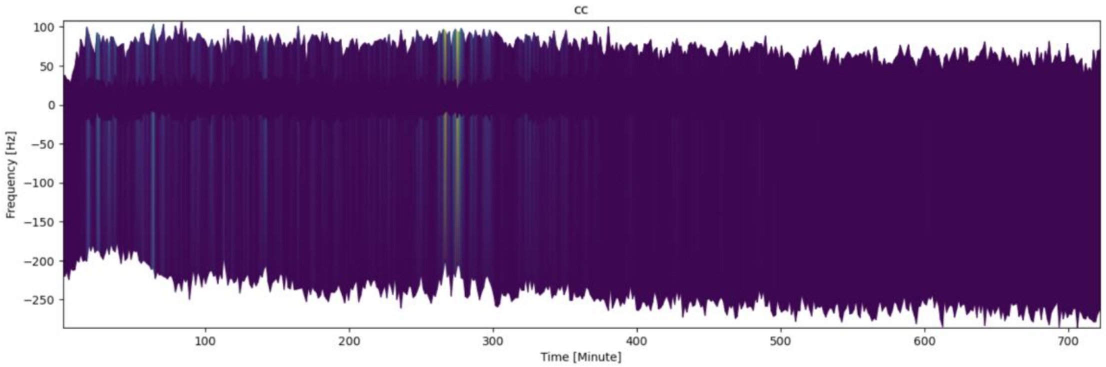
Spectrograms

- A spectrogram is a visual representation of a signal's power over time at various frequencies present in a particular waveform.
- Used to examine the frequency content of a continuous signal recorded by individual stations



Cepstral Coefficients

- Serves as a tool to investigate periodic structures in frequency spectra.
- Computed via the discrete cosine transform of the logarithm of the estimated signal's (power) spectrum.



Wavelet Transform

- Analyzes a signal at different frequencies with different resolutions
  - At high frequencies: good time resolution, but relatively poor frequency resolution.
  - At low frequencies: good frequency resolution, but relatively poor time resolution.
- Shows excellent advantages for the analysis of transient signal(spike and action potentials).
- However, it is not translation invariant, as it computes convolutions with complex wavelets.

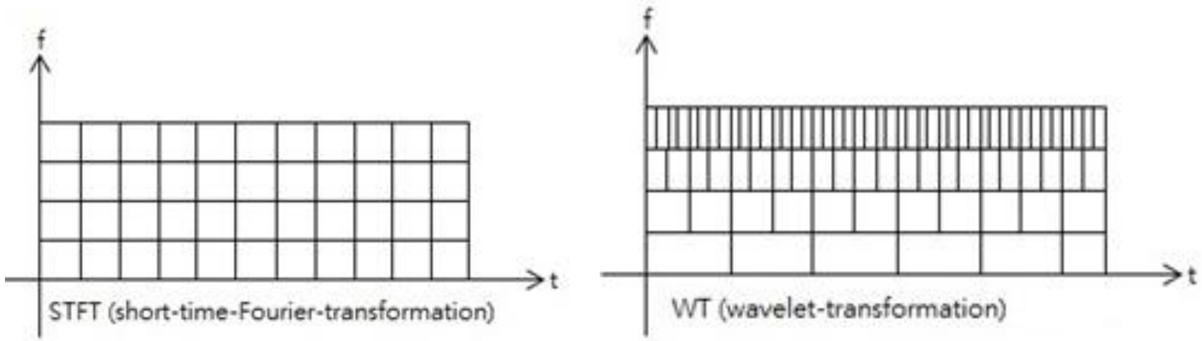
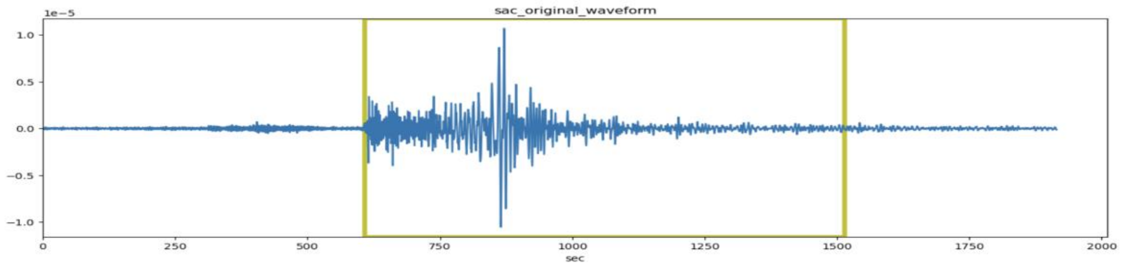
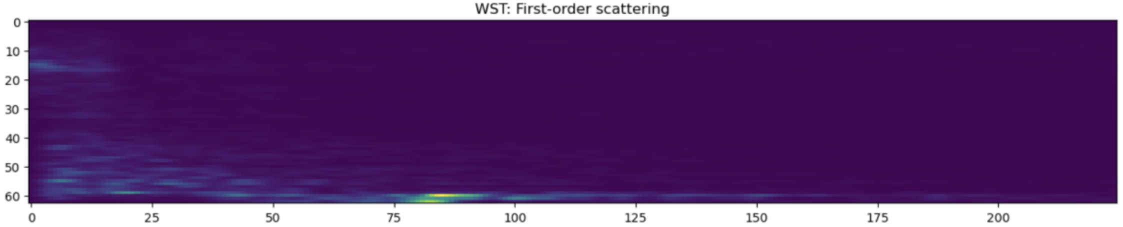
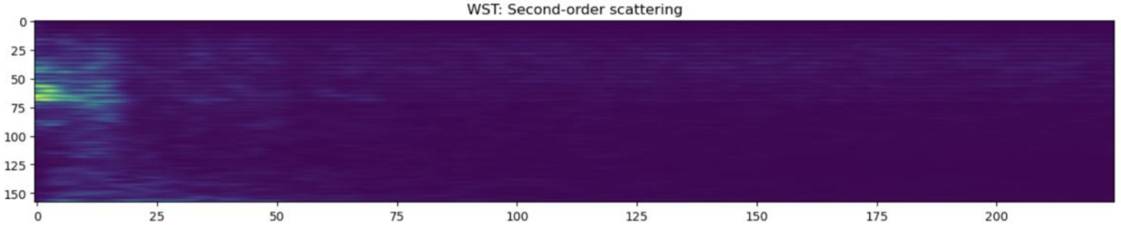


Image obtained from https://en.wikipedia.org/wiki/File:STFT_and_WT.jpg







Slonogram vs Spectrogram

- **Disadvantages of time-based frequency transformations**
  - To ensure consistent length, traces must be cropped.
  - S-wave and P-wave energy travel at different velocities, higher amplitudes will appear in different positions within the spectrogram/cepstral coefficients/ wavelet scattering transform, dependent on epicentral distances to capture the dynamics of earthquakes/explosions

- **Slowness:** It is the inverse of velocity: time (sec) taken for a packet of seismic energy to travel from source to receiver, divided by the epicentral distance (km). The abscissa of the slowness waveform is in units of sec/km.

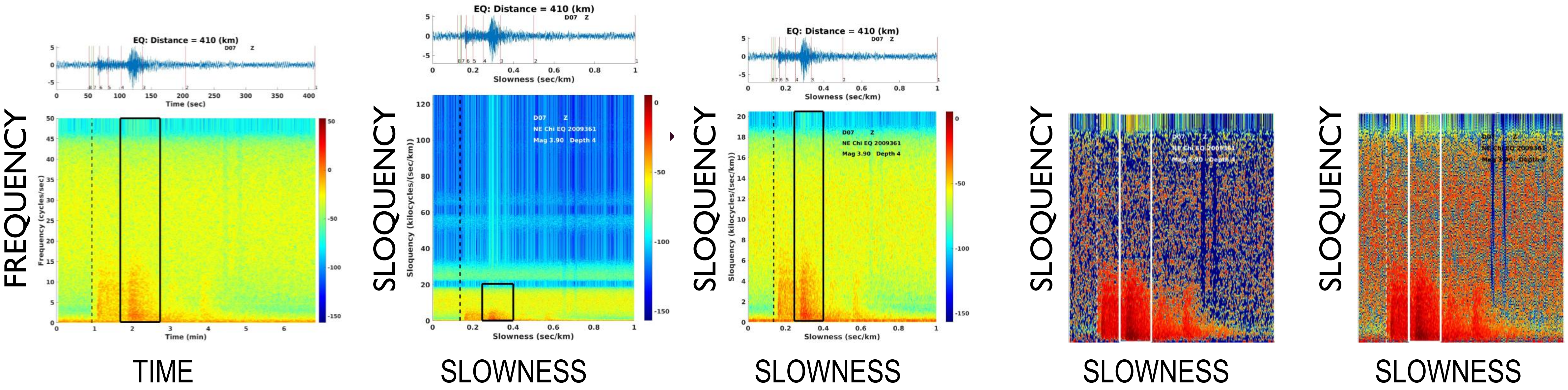
- **To remedy this, we introduce a novel feature transformation method called Slonograms derived from slowness as opposed to time**
  - The position of S and P-energy will be confined to different ranges of slowness.
  - For epicentral distances > 1 km, the slowness waveform will be compressed (not distorted) compared to the time series waveform.
  - The advantage of the slonogram is that S-type energy will appear in nearly the same slowness range, regardless of the source-receiver distance. The same is true for P-type energy only the range will be different.

DEFINITIONS & UNITS:

Time: sec; Frequency: cycles/sec; Slowness: sec/km; Sloquency: cycles/(sec/km)
Δt = sample spacing in time; Δs = sampling spacing in slowness

Pictorial Steps to go from Spectrogram to Slonogram

SPECTOGRAM **SLONOGRAMS**



STEP 1: Create spectogram. Δt defined. Range in x-axis is distance dependent. Frequency range is 0 to Nyquist.

STEP 2: Create slonogram. Δs imposed. Range in x-axis fixed. Imposed sloquency values fixed for all stations.

STEP 3: Top part of slonogram cut. Range in sloquency is 0 to true Nyquist. Number of sloquency values vary by station.

STEP 4: Slonogram optionally denoised. Range in sloquency is 0 to true Nyquist. Number of sloquency values vary by station.

STEP 5: Columns of slonogram interpolated so that there is the same number of rows from 0 to true Nyquist.

Machine Learning Classifiers

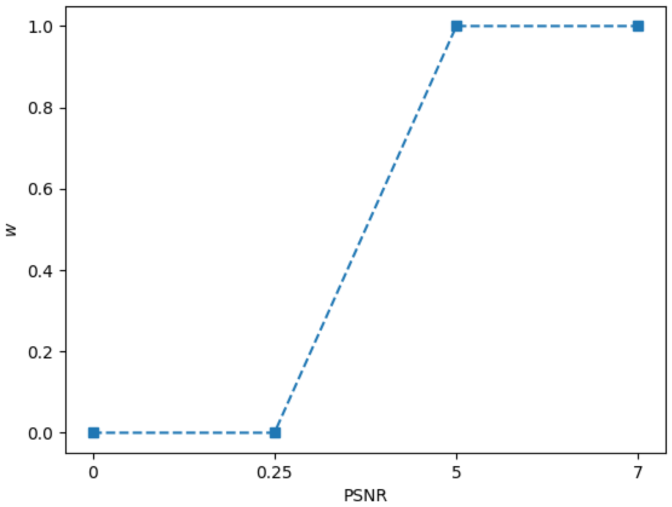
Model	Description	Discriminant	Hyperparameter
K-Nearest Neighbors (K-NN)	This is an instance-based (hence, non-parametric) classification algorithm	Fraction of explosion samples for the given neighbors	Number of neighbors: $N \in \{1, 2, \ldots, 13\}$
Support Vector Machines (SVM)	The core idea of SVM is to find a hyperplane so that distance from it to the support vectors on each side is maximized	Decision function: distance of the sample to the separating hyperplane	Kernel choice: $\mathbf{kernel} \in \{'rbf', 'poly', 'linear'\}$ Regularization parameter $C \in \{1, 2, 3, \ldots, 50\}$
Multi-Layered Perceptron (MLP)	An MLP is a fully connected class of feedforward artificial neural network.	Output of the neural network before applying the argmax operator	Number of hidden_layers, batch size, learning rate, hidden layer dimensionality, selu activation function, probability of dropout.

Network-based event discrimination

- A varying number of stations (the network) record the same event.
  - Stations/traces may be excluded from the network depending on factors such as prevalence of noise.
- The same, pre-trained discrimination pipeline is employed to classify the recorded trace of each station individually (vertical or 3 channel).
- There may be disagreement in terms of labels between such assessments.

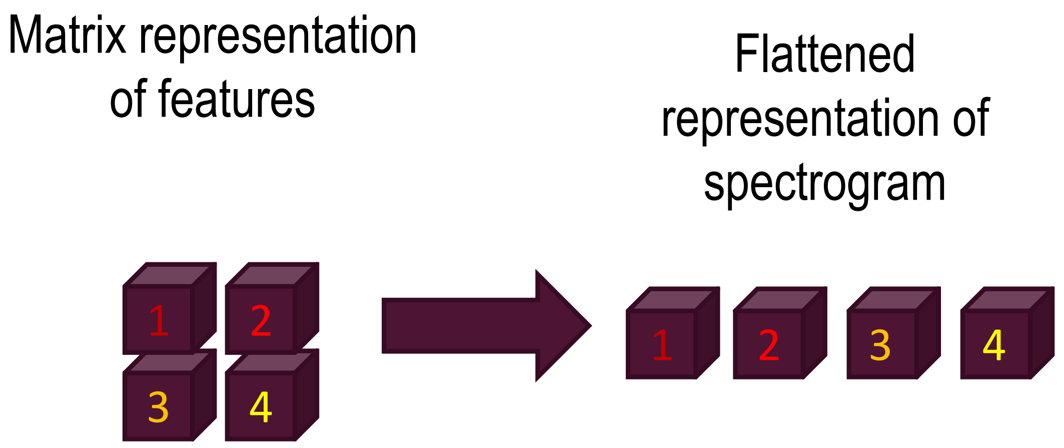
$$w = w_{\max}\ \min\left\{1, \max\left\{0, \frac{\mathrm{PSNR} - \mathrm{PSNR}_{\mathrm{lo}}}{\mathrm{PSNR}_{\mathrm{hi}} - \mathrm{PSNR}_{\mathrm{lo}}}\right\}\right\}$$

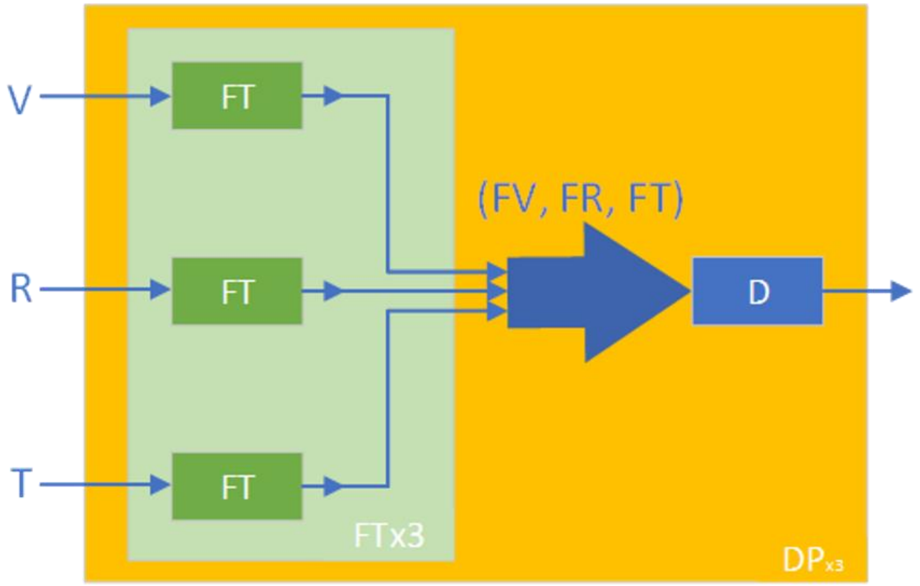
Employed weights for each station as shown below with w_{max}=1, $PSNR_{lo}$= 0.25 and $PSNR_{hi}$=5.



PSNR stands for P-wave SNR, it indicates the strength of the identified P-wave in comparison to a noise sample from the same trace.

How Features are Passed to ML Models





V, R and T stand for the vertical, radial and transverse components of trace respectively, while FT stands for a feature transformation and D stands for a discriminator (classifier), which may output "hard" or "soft" labels. FV, FR and FT are the features obtained from FT, which are concatenated to (FV, FR, FT).

Network-based Event Discrimination

<u>Dataset details</u>

- Maximum duration of traces: 12.05 minutes
- Pre-P wave start time: 10 seconds

<u>Event Splitting</u>

- Training: 64 earthquakes, 40 explosions
- Validation: 8 earthquakes, 5 explosions
- Test: 8 earthquakes, 5 explosions

<u>Hyper-Parameter (HP) tuning</u>

- Tree-based Parzen Estimator (Optuna)
- HP search space is a combination of discrete and continuous search spaces.

<u>Station-wise Accuracy</u>

- Computed across all traces (station-wise) for ***all*** events
- Disregards individual events (class labels).
- $Acc = \frac{Number\ of\ correct\ predictions}{Number\ of\ traces}$

<u>Network Accuracy</u>

- Computed across all traces ***per*** event.
- Will only represent one class (associated with the event)
- $Acc = \frac{Number\ of\ correct\ predictions\ for\ event}{Number\ of\ traces\ for\ event}$

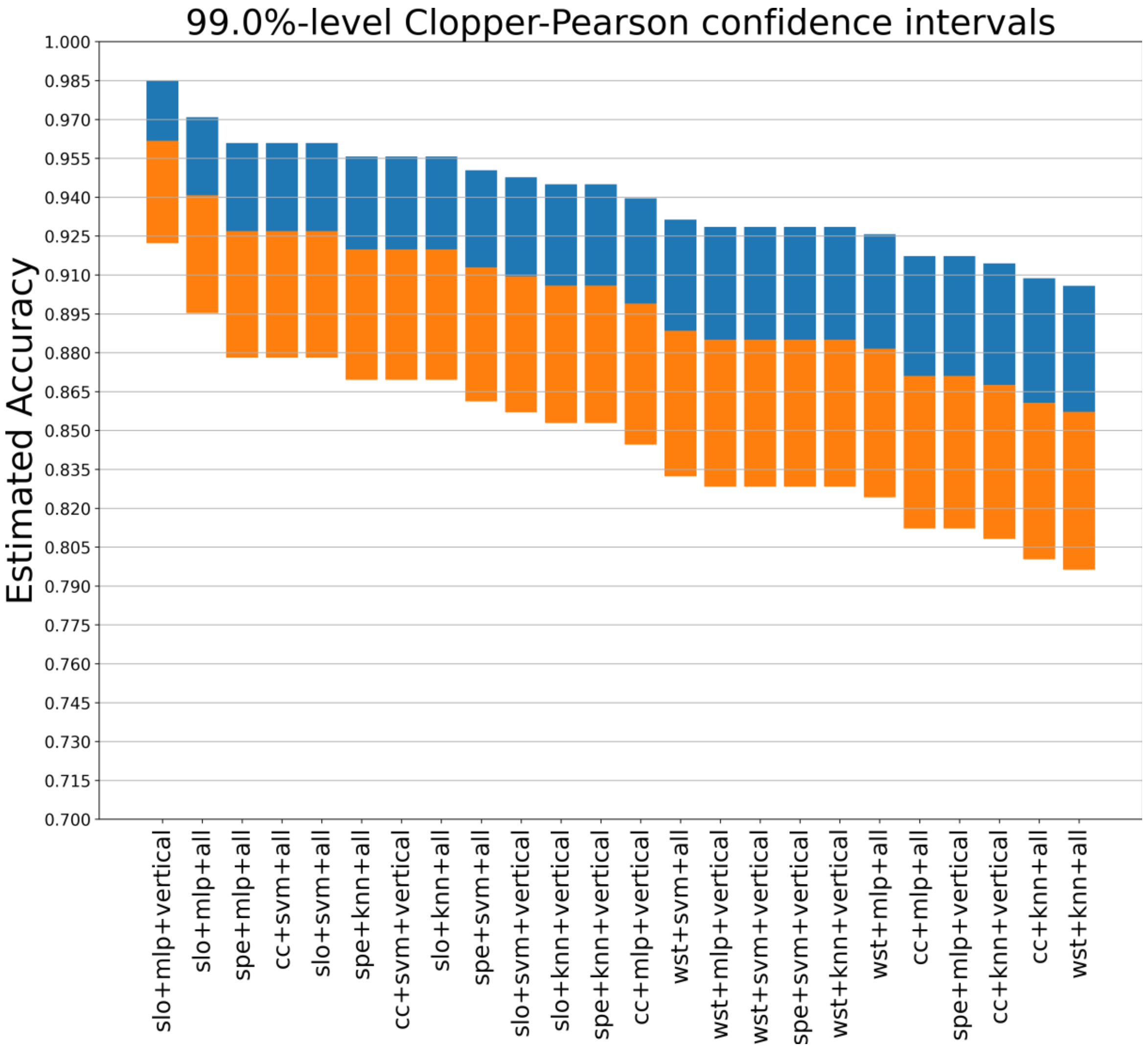
Results: Station-based Discrimination Performance

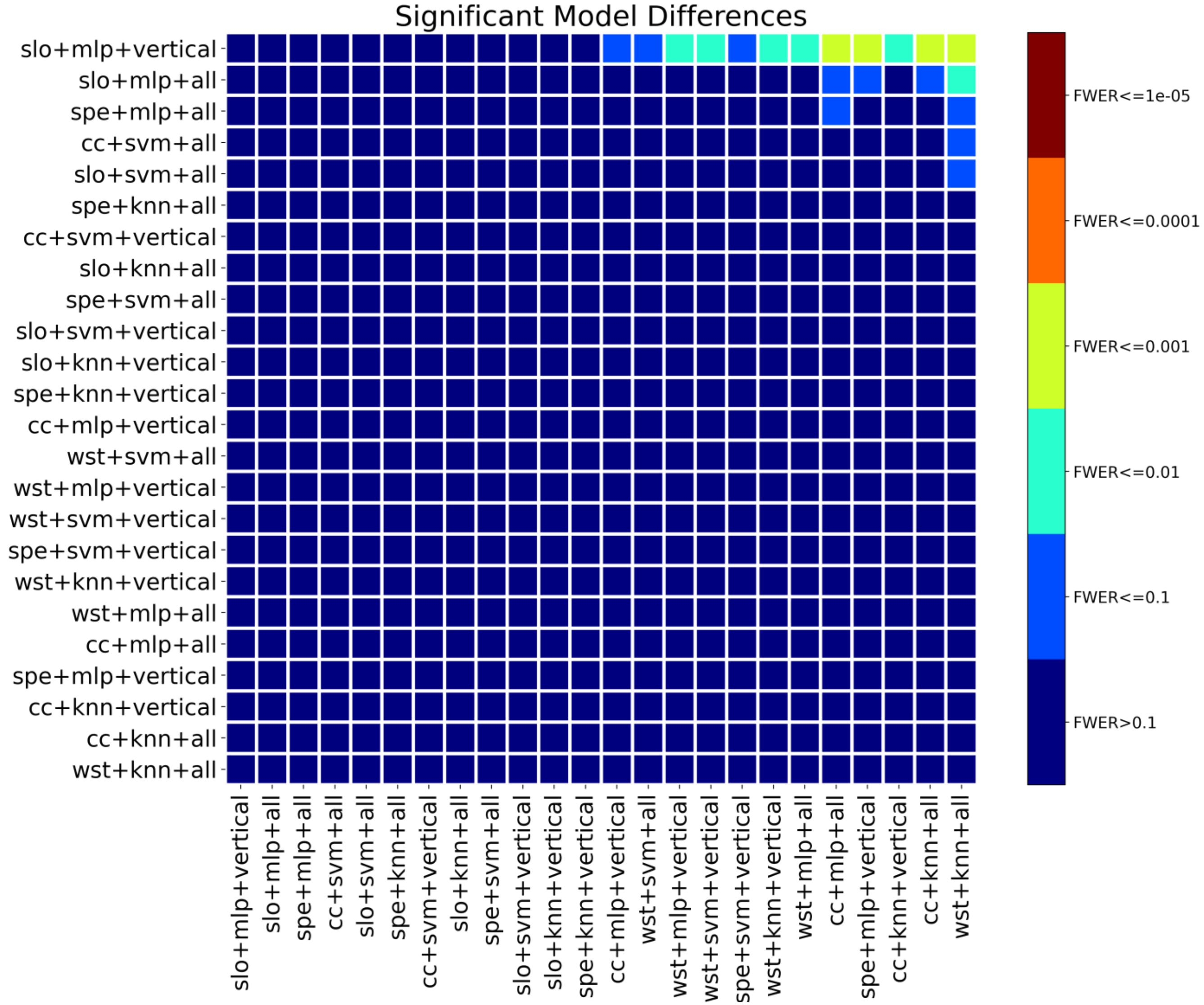
Feature Extraction method + Model	Station-based accuracy (vertical channel)	Station-based accuracy (3 channel)
Slonogram + MLP	**96.16**	**94.08**
Slonogram + SVM	90.94	92.68
Slonogram + KNN	90.59	91.99
Spectrogram + SVM	88.51	91.29
Spectrogram + KNN	90.59	91.99
WST + KNN	88.51	85.71
CC + SVM	91.99	92.68
WST + SVM	88.51	88.85
WST + MLP	88.51	88.15
CC + KNN	87.76	86.06
CC + MLP	89.9	87.1
Spectrogram + MLP	87.11	92.68
Naïve	88.51	88.51

Test dataset

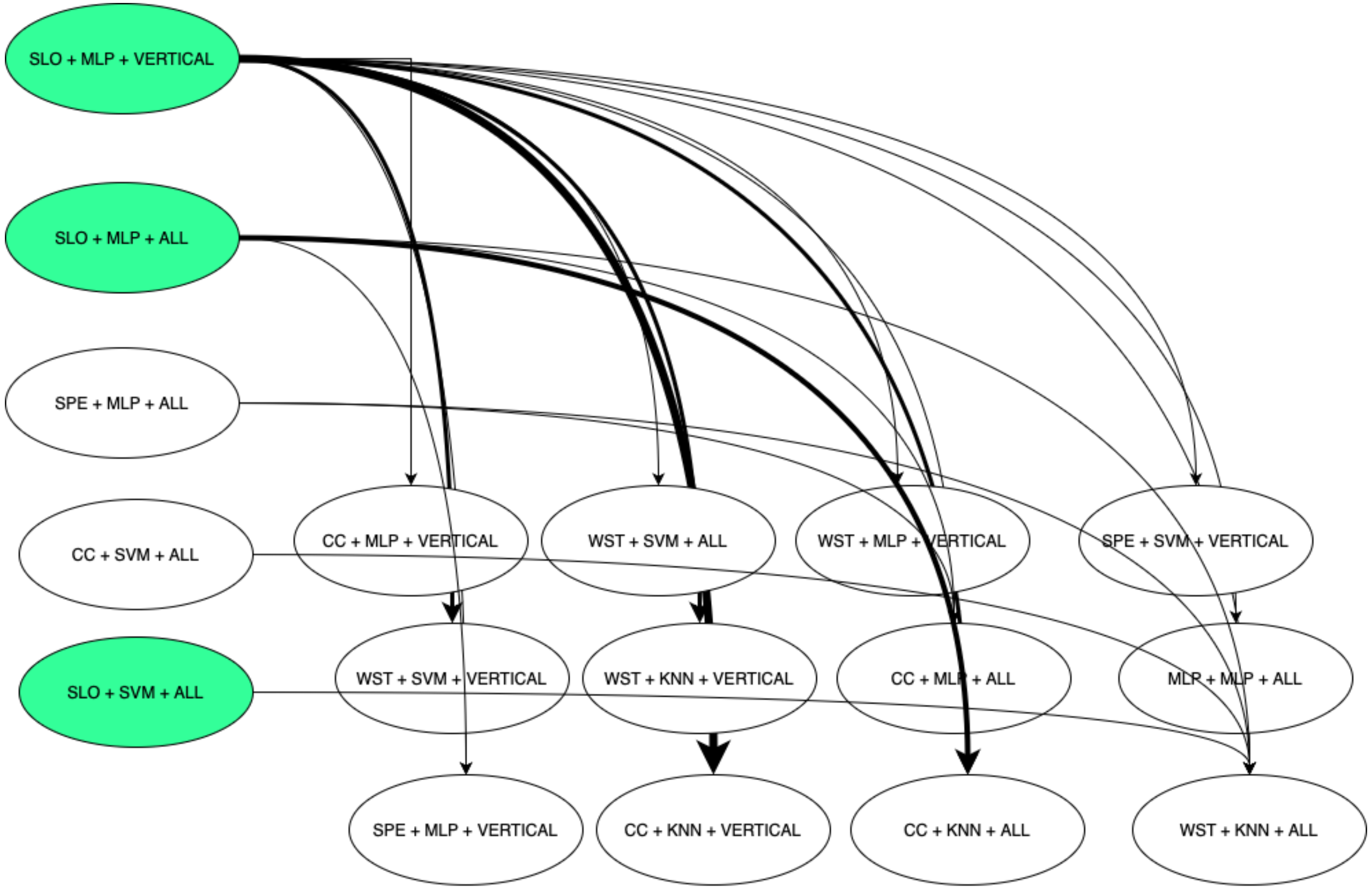
- 254 earthquakes and 33 explosions traces.
- For both vertical and 3-channel results, the same number of hyper-parameter search trials were conducted.
- Slonograms have no hyper-parameters, i.e., parameters remain fixed for interpolation algorithm and spectrogram.

Results: Station-based Discrimination Performance





Results: Station-based Discrimination Performance



Significant differences

- 3 out of the 5 dominating pipeline use Slonograms (highlighted in green).
- Slonogram + MLP + vertical pipelins is significantly better than the last 12 performing models (out of 24 models)
- Edge thickness implies a significance with a smaller Family-Wise Error Rate (FWER)

Results: Network-based Discrimination Performance

- **Event-based Discrimination Performance**
  - Test dataset has **8 earthquakes and 5 explosions**
  - CC + SVM is the best model for the chosen discriminant function.
  - 3 channel data

Feature Extraction method + Model	Network accuracy (vertical channel)	Network accuracy (3 channel)
Slonogram + MLP	69.23	69.23
Slonogram + SVM	76.92	76.92
Slonogram + KNN	61.54	84.61
Spectrogram + SVM	76.92	69.23
Spectrogram + KNN	69.23	92.31
WST + KNN	61.53	61.54
CC + SVM	**92.31**	**92.31**
WST + SVM	61.53	61.54
WST + MLP	61.53	69.23
CC + KNN	76.92	76.92
CC + MLP	69.23	69.23
Spectrogram + MLP	69.23	69.23
Naïve	61.54	61.54

Summary

- We first constructed several machine learning pipelines for discrimination of earthquakes and nuclear explosions.
- Furthermore, we introduced a geophysics-inspired feature transformation technique – slonograms – to account for epicentral distances that showed the best performance for station-based accuracies.
- Station-based performance did not translate into network-based performance, necessitating another look at how to better gather consensus among individual predictions.
- We notice that considering all 3 channels did not significantly change the discrimination performance.