T2.1-453

Global and Regional Explosion Discrimination: Improving Machine Learning Models with Seismology-Driven Features

Louisa Barama and Qingkai Kong

Lawrence Livermore National Laboratory

INTRODUCTION AND MAIN RESULTS

Accurately distinguishing between natural and anthropogenic (explosive) seismic events is crucial for global and regional monitoring efforts. Recent advances in machine learning, particularly when enhanced with seismology-led features, offer promising improvements in explosion discrimination models. Our study demonstrates that current models achieve $\geq$ 90% accuracy in distinguishing explosions from other seismic events. While adding new seismological features can increase the robustness of these machine learning models, the observed improvements have so far been incremental. Importantly, even with limited availability, legacy seismic data continues to provide valuable insight into model performance and the detection of underground explosions.

This Low Yield Nuclear Monitoring (LYNM) research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). The authors acknowledge important interdisciplinary collaboration with scientists and engineers from LANL, LLNL, NNSS, PNNL, and SNL.
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC.

SnT 2025 — 8 SEPTEMBER ONLINE DAY — 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

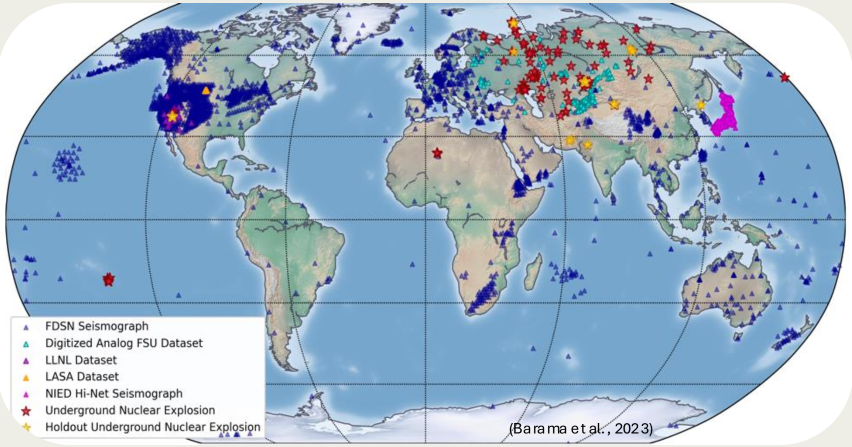
Global and Regional Explosion Discrimination: Improving Machine Learning Models with Seismology-Driven Features

Louisa Barama and Qingkai Kong

T2.1-453

Introduction

While seismic explosion data is limited compared to natural earthquakes, it has proven useful for labeling data in machine learning (ML) applications. Recent work by Barama et al. (2023) and Kong (2022) has advanced our ability to discriminate between earthquakes and explosions using both digital and historical seismograms. Building on these efforts, we explore how ML models can be further developed to improve explosion discrimination at both regional and teleseismic distances.

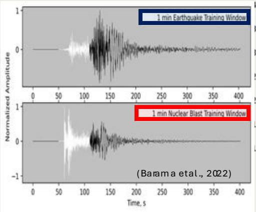


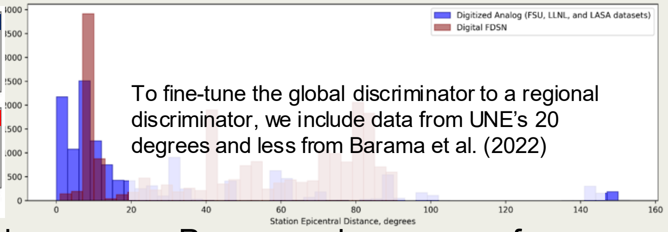
Map showing the location of seismic stations (triangles) and source Underground Nuclear Explosion (UNE) (red stars). Color of triangle denotes dataset source institution.

The goals are to assess the transportability of the global model from Barama et al. (2023) fine-tuned to regional distances, make use of legacy UNE data, and evaluate the impact of adding features to training. Through this research, we aim to demonstrate that machine-learning classifiers can be robust tools for seismic monitoring and can provide valuable insight into the factors that influence model predictions for explosive events.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Data and Methods





To fine-tune the global discriminator to a regional discriminator, we include data from UNE's 20 degrees and less from Barama et al. (2022)

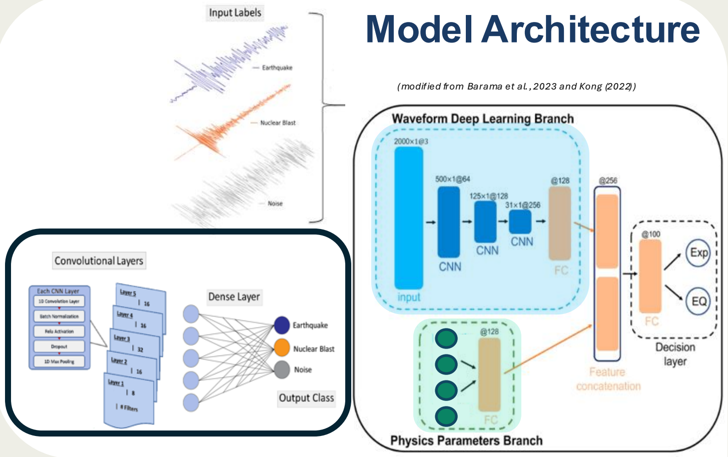
In our approach, we use P-wave seismograms from both digital and formerly-analog records, extracting a range of derivative features such as radiated earthquake energy, polarity, dominant frequency, coda wave envelopes, and phase ratios.

- Limitations:
  - Many repeated test locations and emplacement conditions for UNEs
  - To address this : training and testing data separated by unique source location / test
- Data Specifications:
  - 5 – 20 epicentral distance degrees
  - P-wave seismograms (20 – 60s)
  - 1 – 5 Hz frequency bands
- Features:
  - Dominant frequency content
  - SNR difference at two frequency bands
  - Polarity
  - P coda decay (slope)

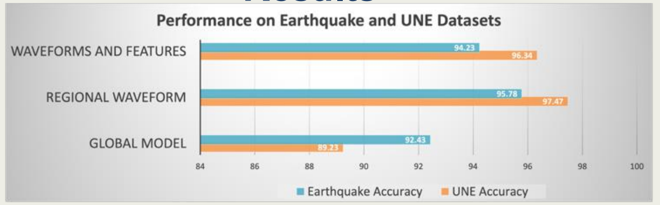
Dataset	Earthquake	Nuclear
Train	25,170	25,170
Test	2500	2500
UNE holdout	1044	1044
Earthquake holdout	1000	1000

*** Data augmentation by randomly varying the P-wave arrival 3 times per waveform for UNE data.**

Model Architecture



Results



- ❖ Achieved model accuracy centering around 90%.
- ❖ Incorporating additional features improves model robustness.
- ❖ Legacy data, despite limited availability, remains valuable for evaluating model performance and underground explosion detection.
- ❖ Regional distance data introduces greater variability than teleseismic data, resulting in slight performance decreases, though models remain robust for new emplacement locations.

On-Going Work

Evaluating the impact of additional component data, such as multiple frequency bands and coda envelopes, to further enhance model performance beyond single frequency band or raw data approaches.

Lawrence Livermore National Laboratory