SAND2021-8363C

Sandia National Laboratories

Exceptional service in the national interest

Can Synthetic Seismic Data Make Up For Observational Sparsity?

Lisa Linville, 06754

Work informed by collaborations with Rob Porritt, Thomas Catanach, Kathy Davenport, and Chris Young

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

TOPIC OVERVIEW

Problem Intro

Synthetic Event Catalogs

Large Stochastic Catalogs

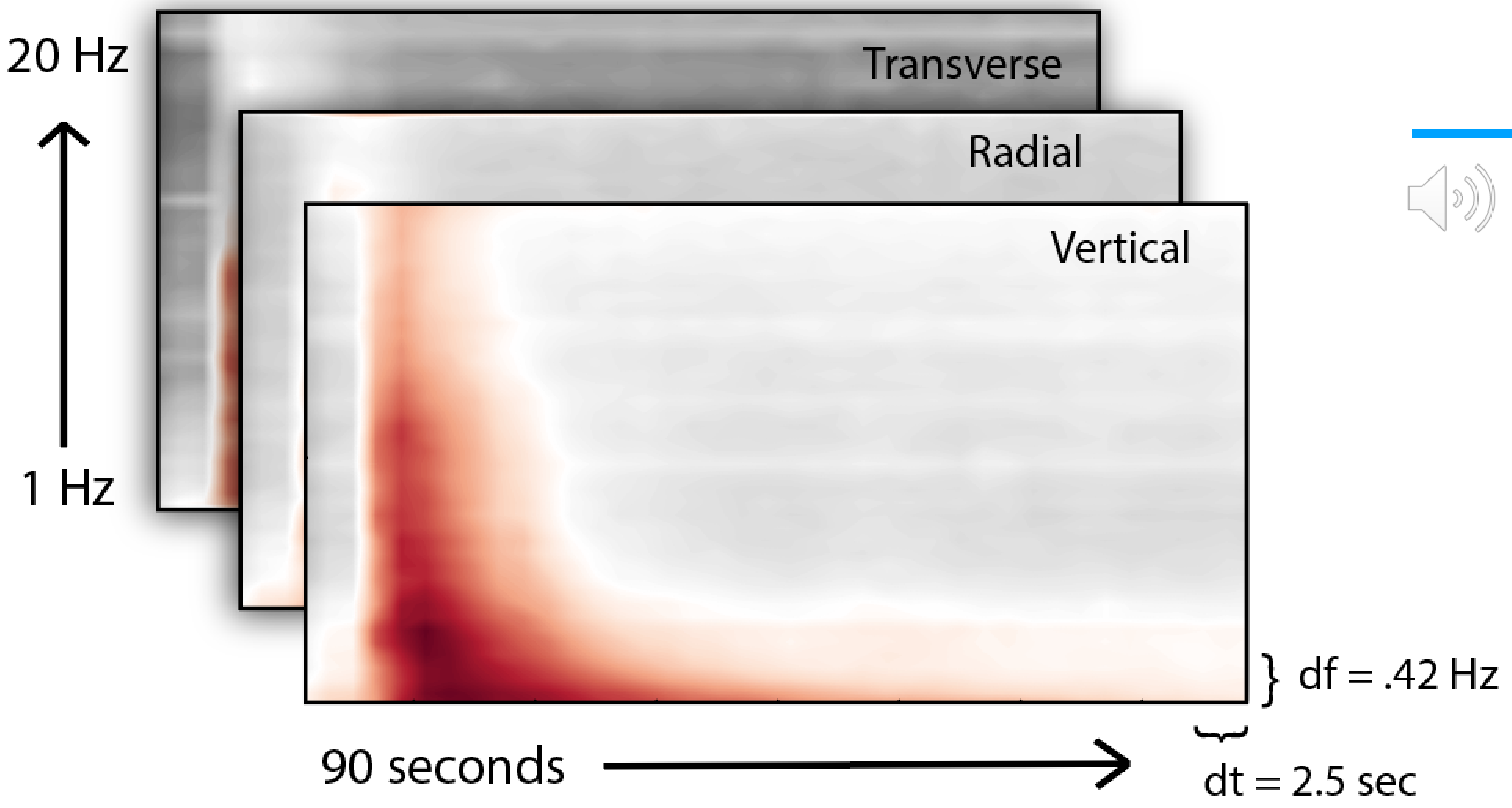
[conditional] GANs

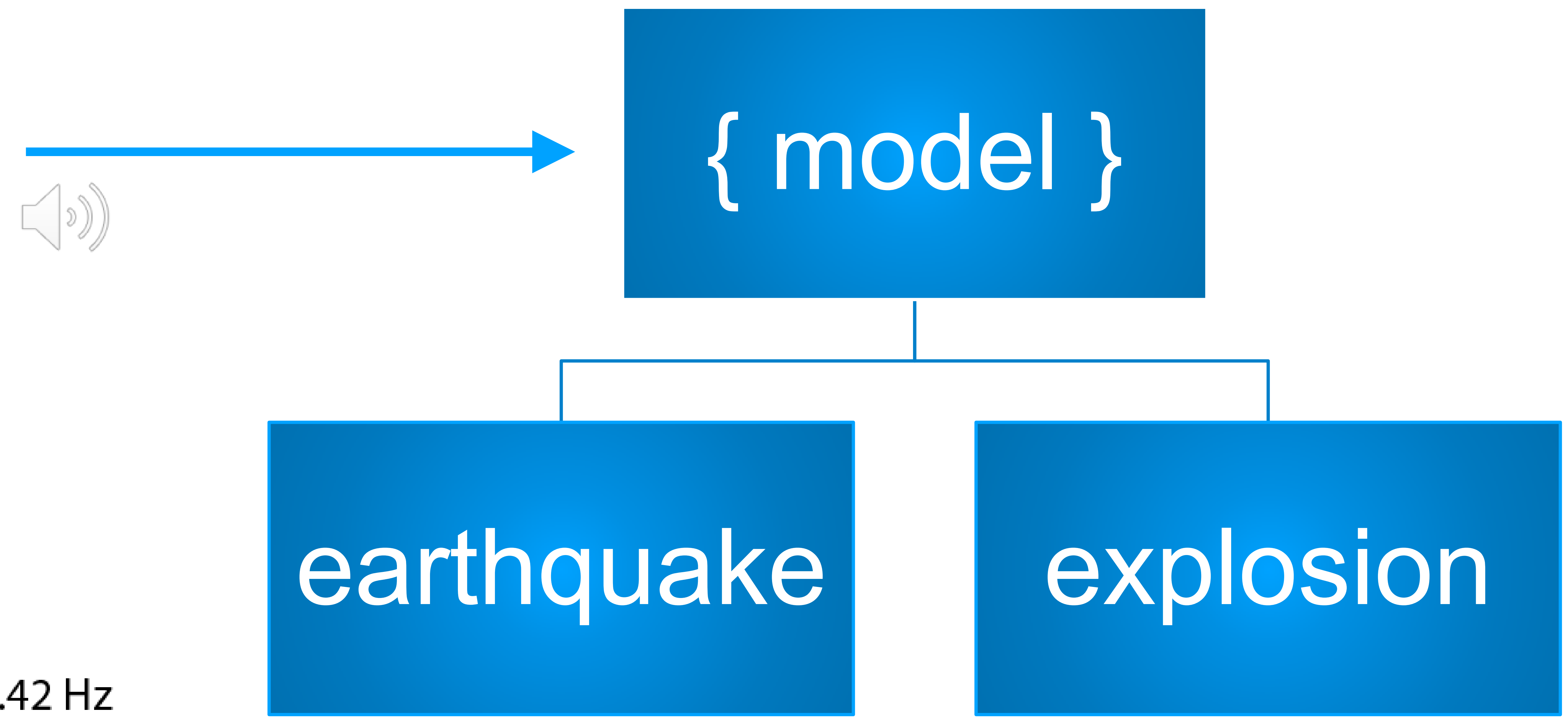
We can demonstrate human level performance *or higher* on a range of common data processing tasks in test studies.

- Quality Control
  - Denoise
  - Gap
  - Glitch
- Detect
  - Arrival vs noise
  - Phase labels
- Associate
  - Temporal concurrence
  - **Competitive assignments**
- Characterize
  - Location
  - Magnitude
  - Type

Focus on using synthetics to improve event type predictions

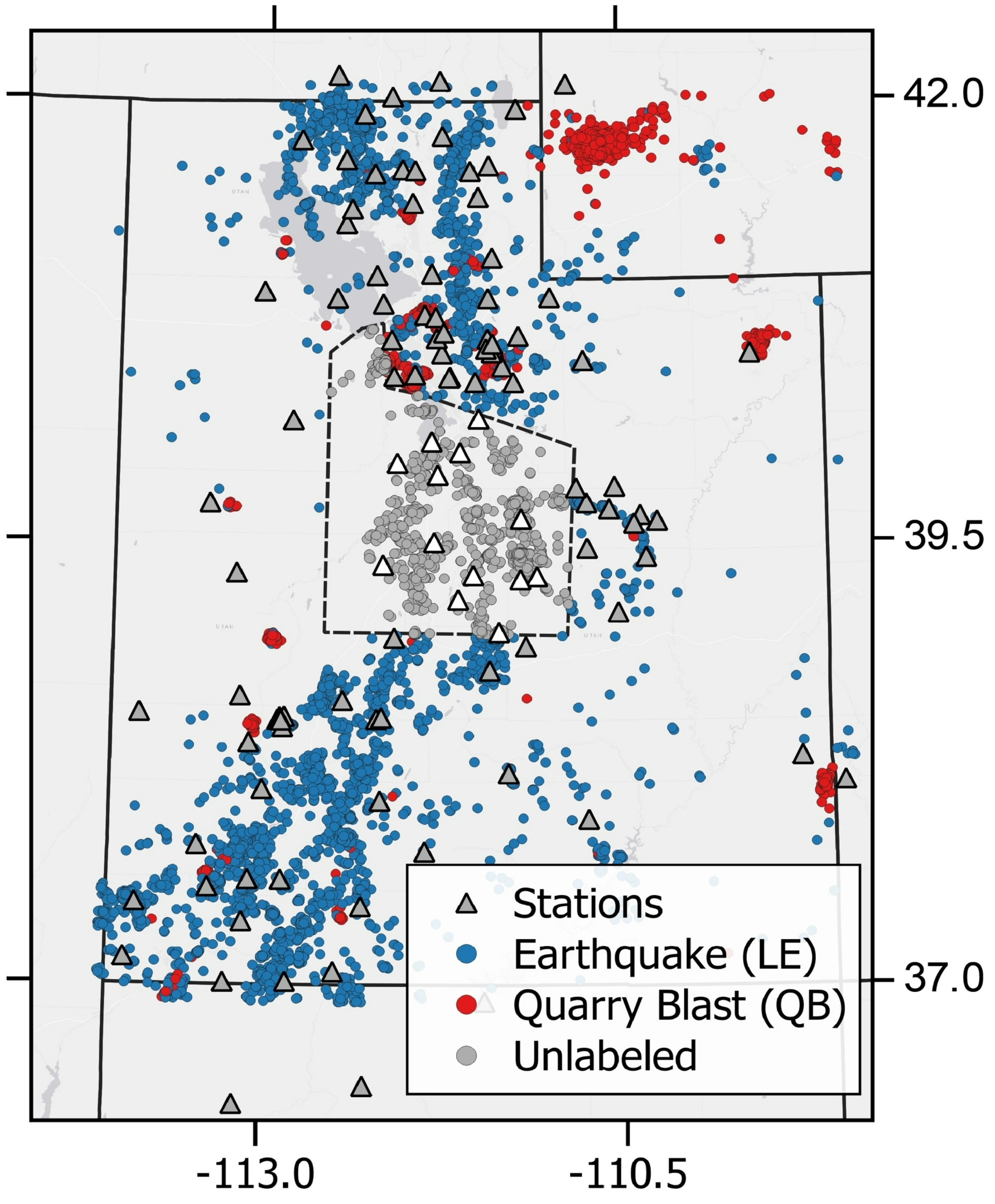
Task: Binary Classification



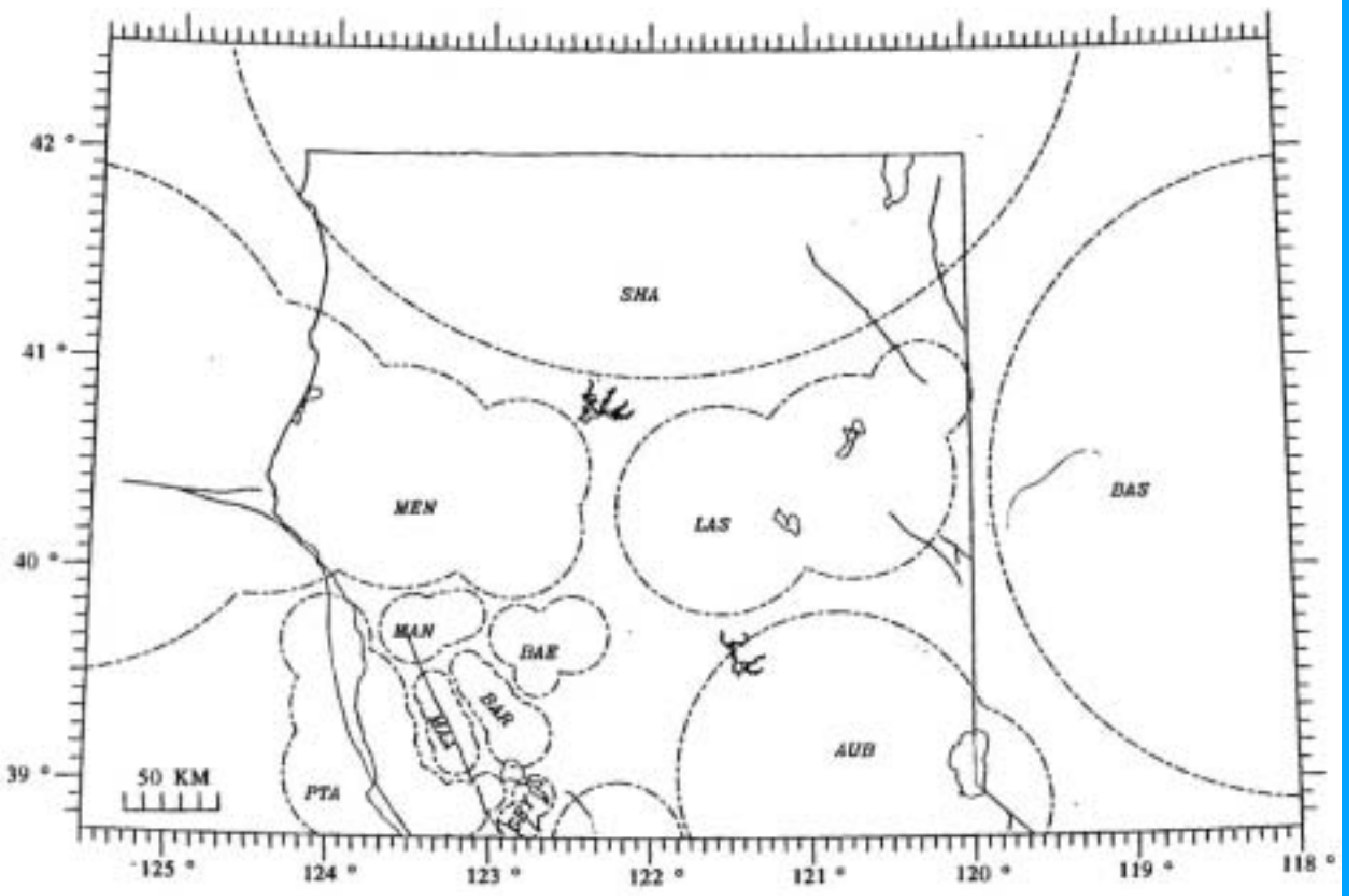


Failure to generalize across subsurface differences

Significant average crustal properties over long distances/scales

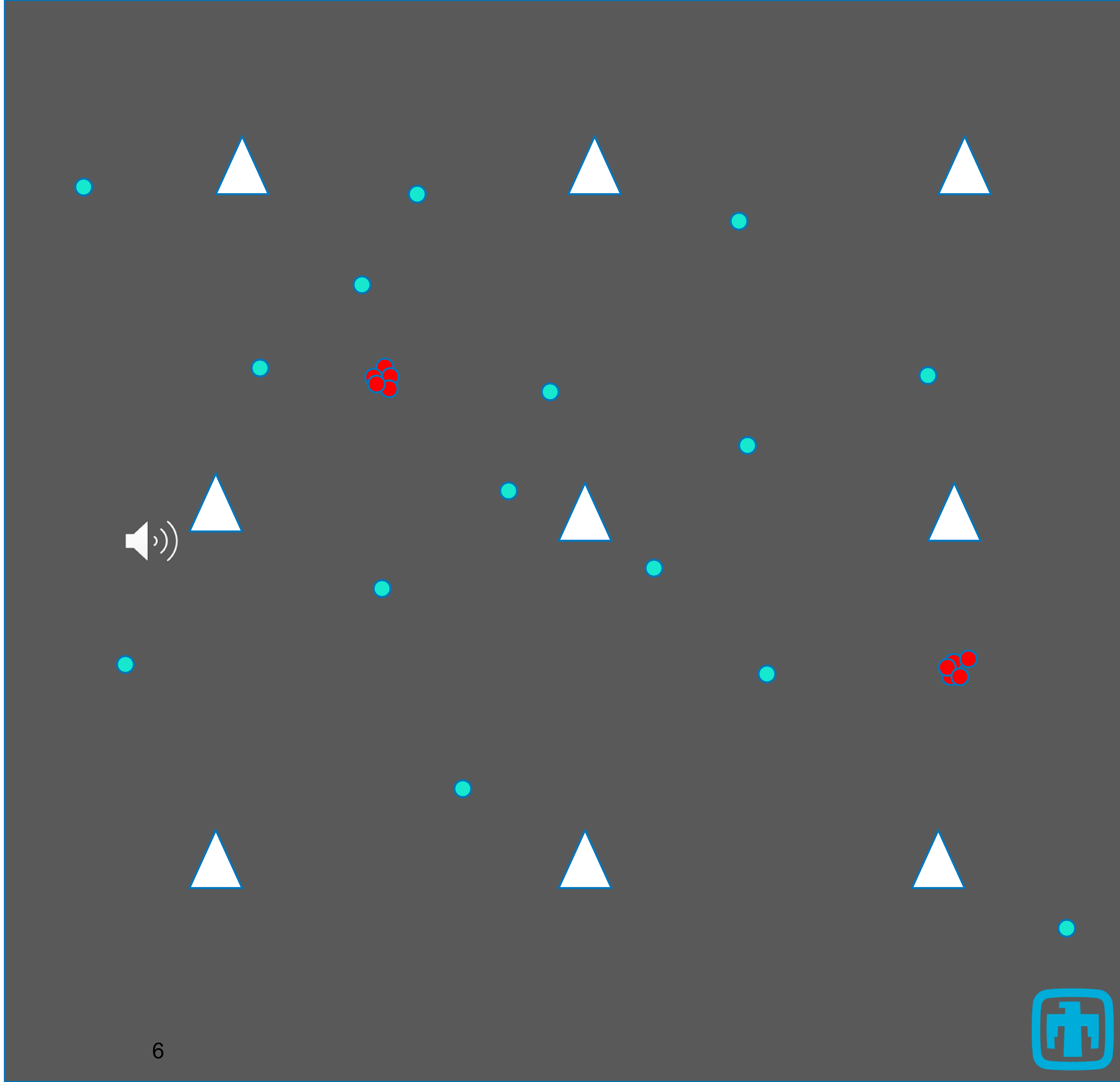


Shallow subsurface, site, or local crustal variation with significant impact on waveform characteristics at higher frequencies

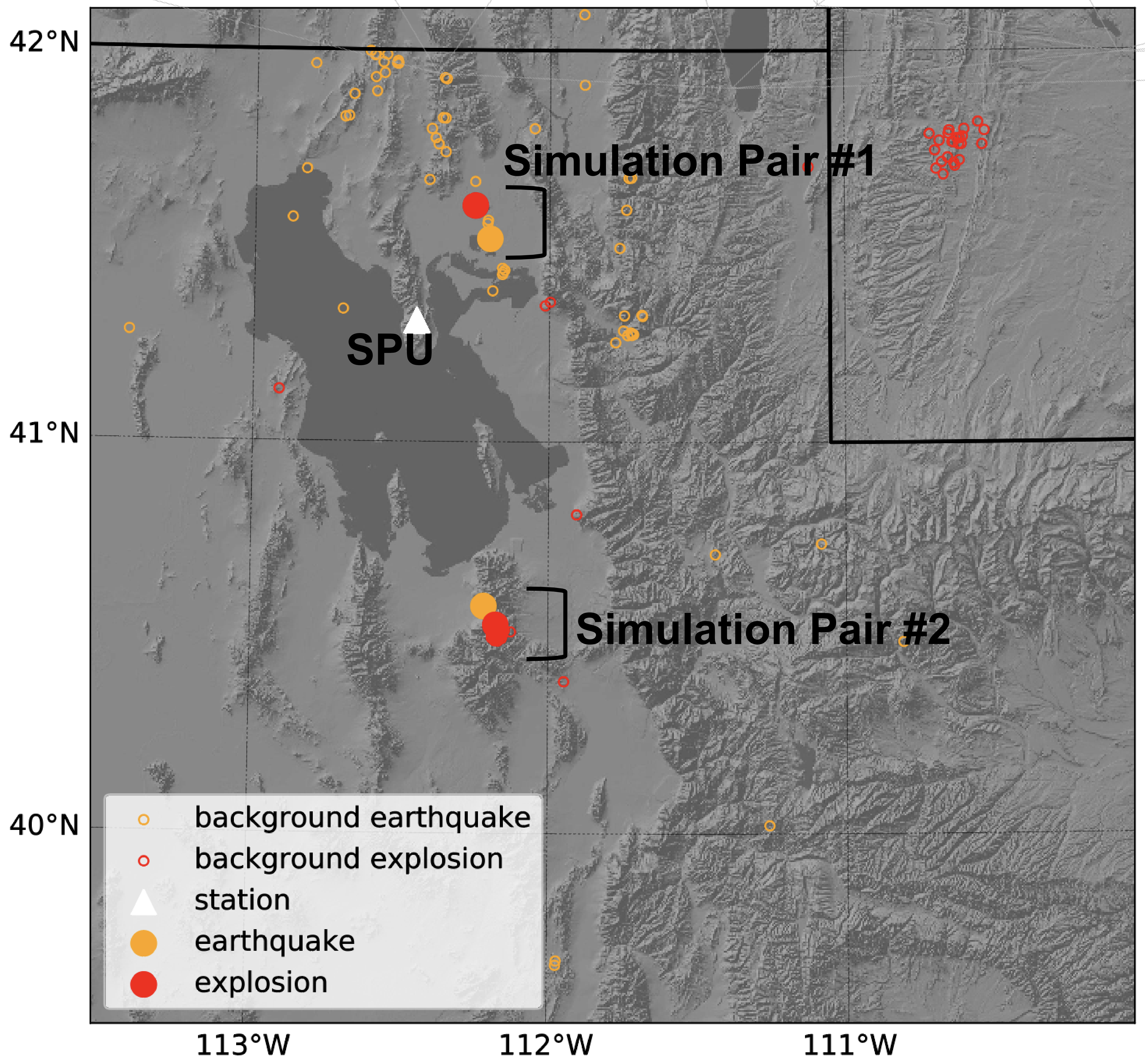


Can we accurately generate synthetic waveforms that match observations with sufficient fidelity to use them for model training?

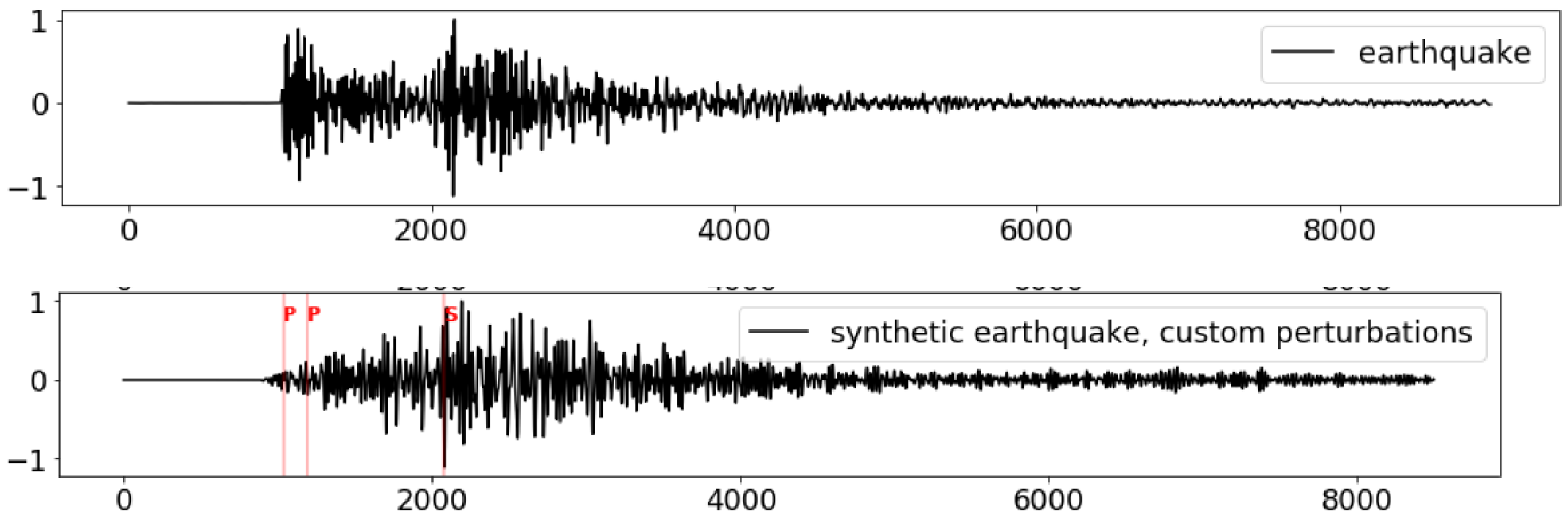
- Rare events
- New locations
- New stations



Generating realistic local – regional synthetics

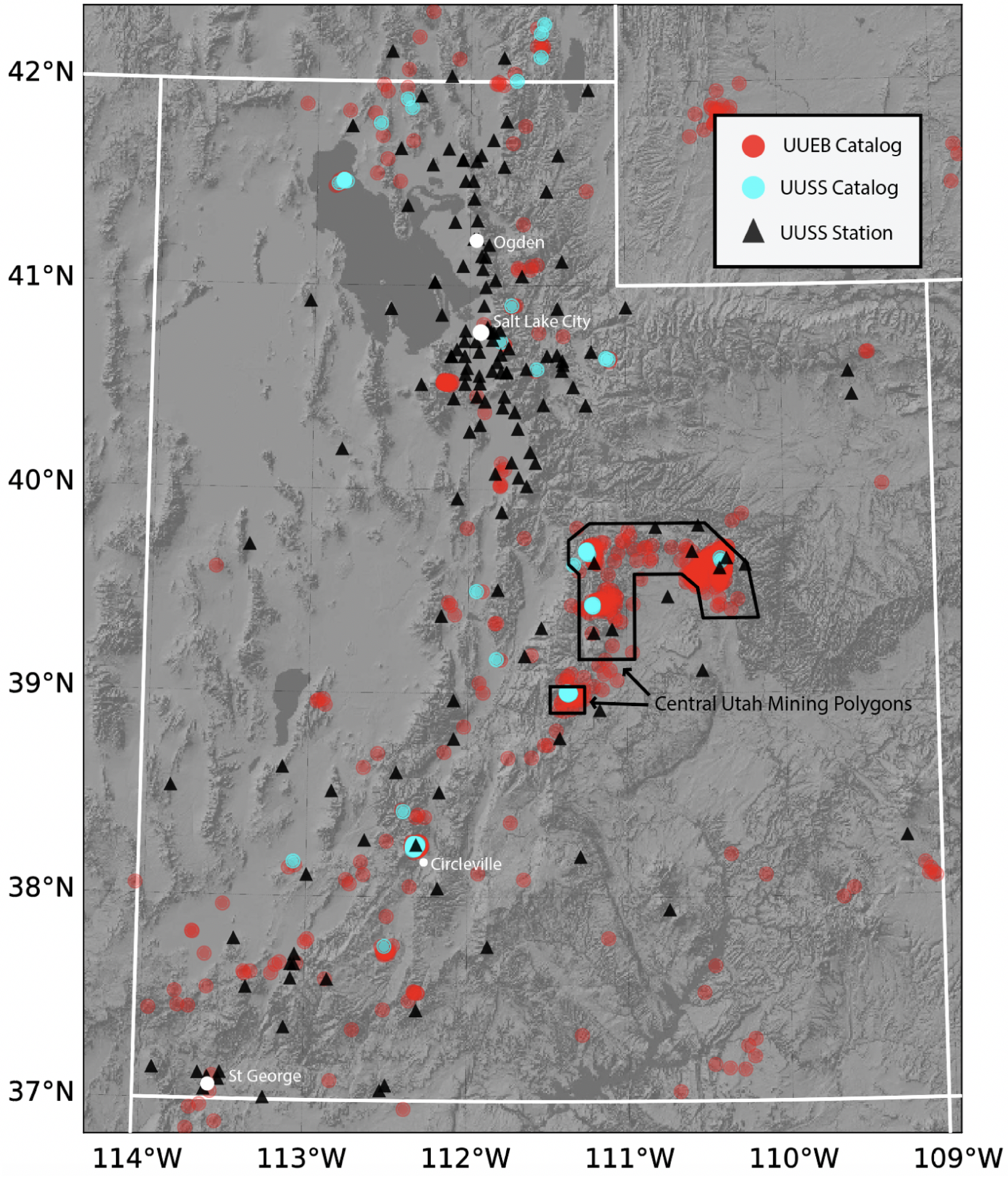


1. Earthquake simulation
2. Explosion simulation
3. Explosion at earthquake depth
4. Find optimal solution



Predictable behavior for new event types?

- 2 weeks of seismicity
- 147 catalog events vs 7k new events
- New source types



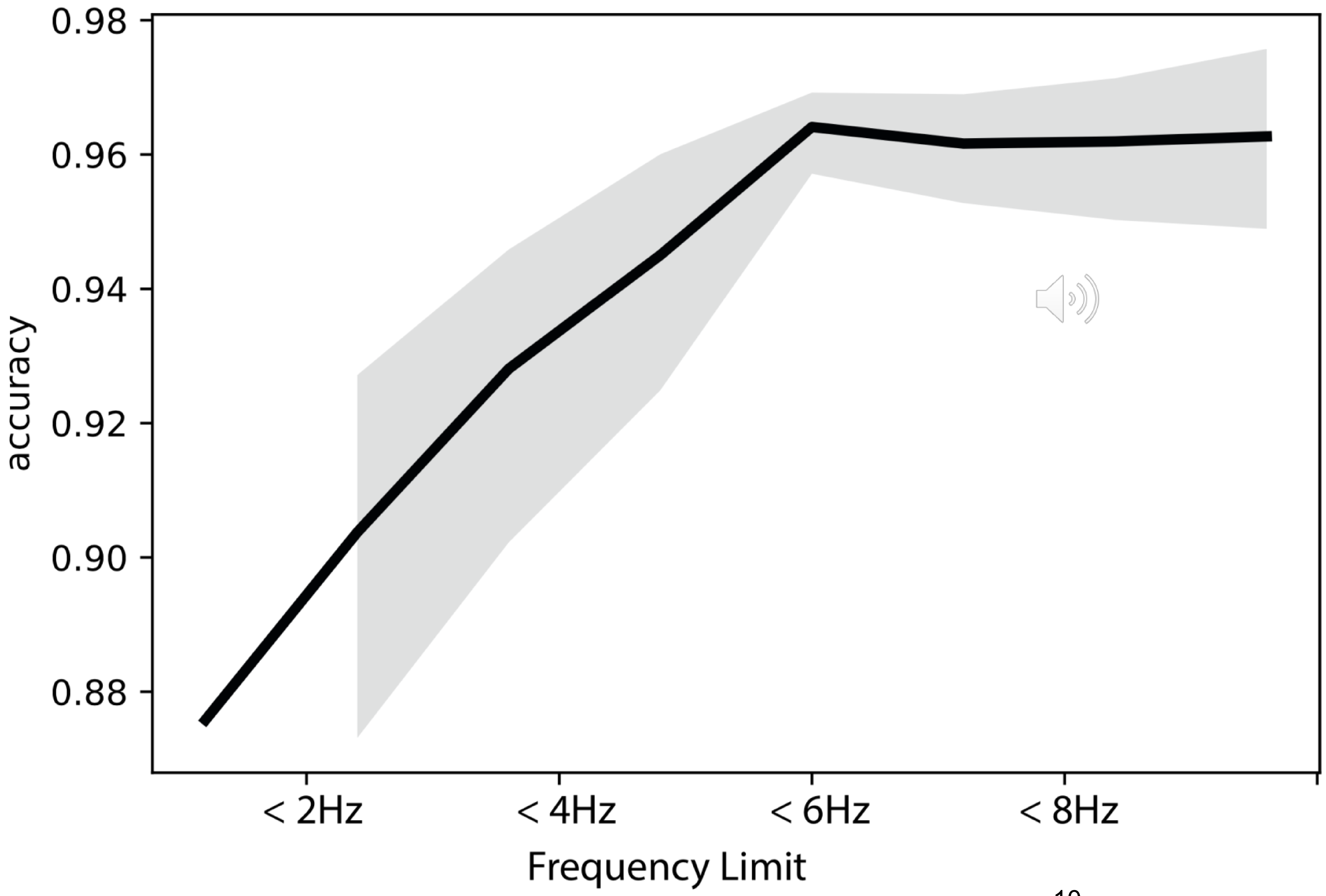
UUSS: University of Utah Seismograph Stations
UUEB: Unconstrained Utah Event Bulletin

More groundwork needed for local waveform fitting

Intractable to optimize fit over the entire catalog

Key challenge: designing metrics that reflect quality of fit according to waveform characteristics of interest

What do ML models need most for effective learning?



Higher frequency waveform characteristics can be harder to model.

The good news is those characteristics might not be as diagnostic.

Stochastic Variations

- Velocity Models
- Source Parameters
- Complexity

How to incorporate real examples?

Will the extreme variation result in meaningful decisions?

Will the issues of scale outweigh the benefit of the model?

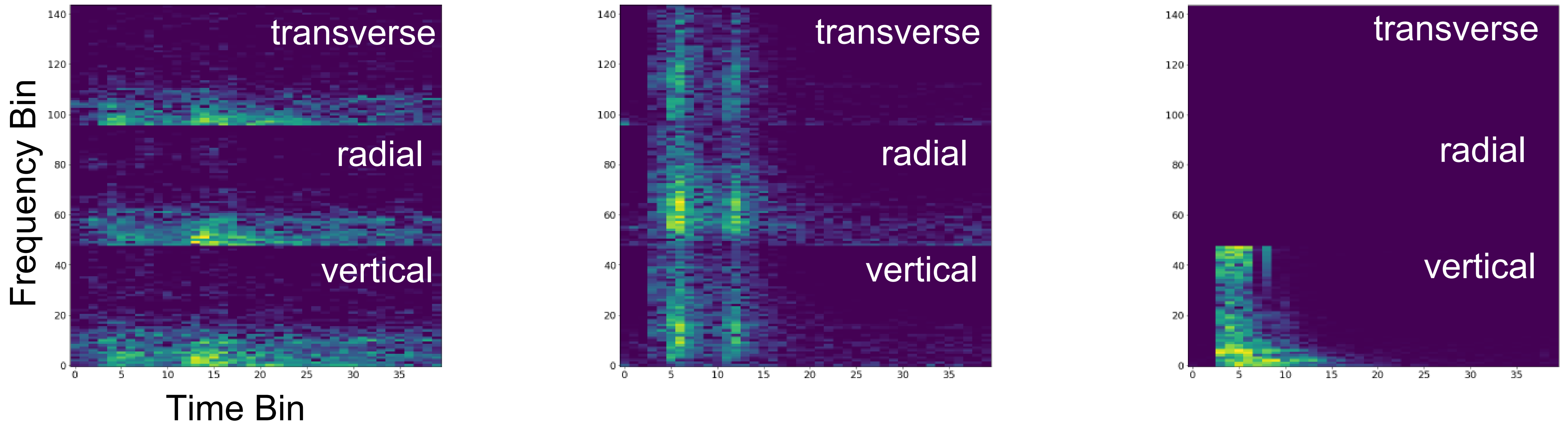
Can Data-Driven Synthetics be useful?

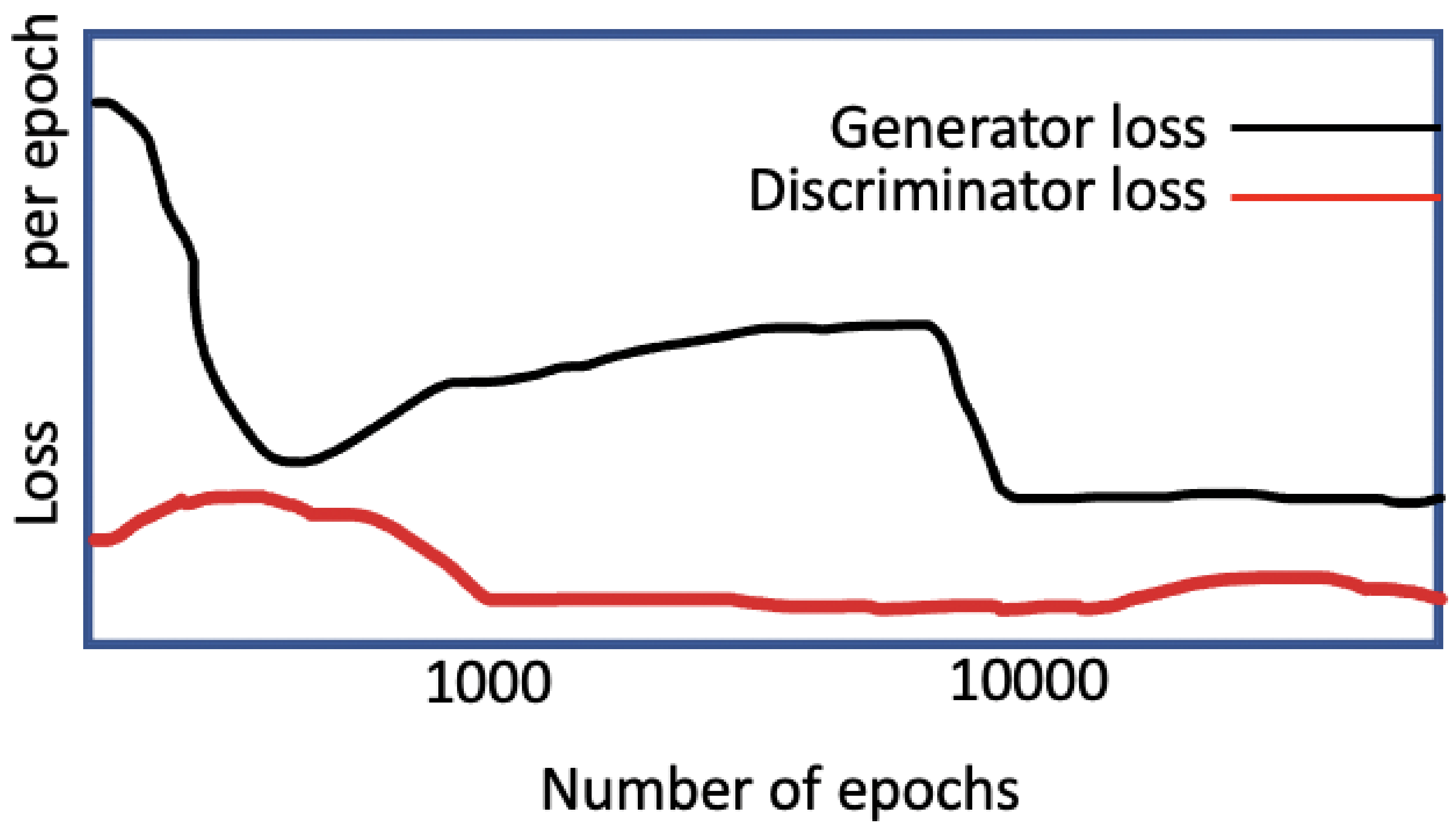
Input: noise vector	Input: data, generator images
Generator: NN with the task of making realistic output	**Discriminator**: NN with the task of identifying fake images
Loss: low if the discriminator can't tell it's a fake	Loss: low if it can tell generator data from real data
Output: n x n x n (image)	Output: binary (decision)

Generative Adversarial Networks (GANs)

- Sample from a learned distribution
- Generate new physically plausible examples according to specific conditions

GANS on Utah Data





Vanilla GAN

they look real but are too unspecific

Conditional GAN*

they can look real but suffer from mode collapse and non-convergence

*conditional gan input: noise + label

GANS on Synthetic Pulses

Conclusions

1. Higher frequency waveform characteristics are hard to model.
2. Currently, good fits are possible but require manual tuning per event.
3. Large catalogs with stochastic variation across unknowns is our current approach.
4. GANs are an interesting idea, but new methods that address current issues need to be explored

Thank You

Questions: llinvil@sandia.gov

This 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, MSTS, PNNL, and SNL

Sandia National Laboratories is a multimission laboratory operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration. Sandia Labs has major research and development responsibilities in nuclear deterrence, global security, defense, energy technologies and economic competitiveness, with main facilities in Albuquerque, New Mexico, and Livermore, California. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.