Sandia National Laboratories

Deep Learning Seismic Signal Detection on the International Monitoring System

Jorge A. Garcia, Stephen Heck, Christopher J. Young

Abstract 1

Producing a complete and accurate set of signal detections is essential for automatically building and characterizing seismic events of interest for nuclear explosion monitoring. Deep learning methods have been applied to local distance earthquake monitoring, demonstrating performance that clearly exceeds current operational methods. We explore the process of training our own iteration of the PhaseNet (Zhu et al., 2019) deep learning model using data from the Provisional Technical Secretariat of the Comprehensive Nuclear-Test-Ban Treaty Organization's International Monitoring System (IMS) global network. We specifically consider only non-array three-component stations of the IMS to ensure the new model is using data that is similar to what was used for the original PhaseNet model.

Deep Learning Model: 2

- PhaseNet is based on U-Net [Ronneberger et al., 2015], originally used for image segmentation.
- Generate 1D segmentations of P, S and noise probabilities
  - Output 3 channels with each being the probability of a data point belonging to the class at each timestep

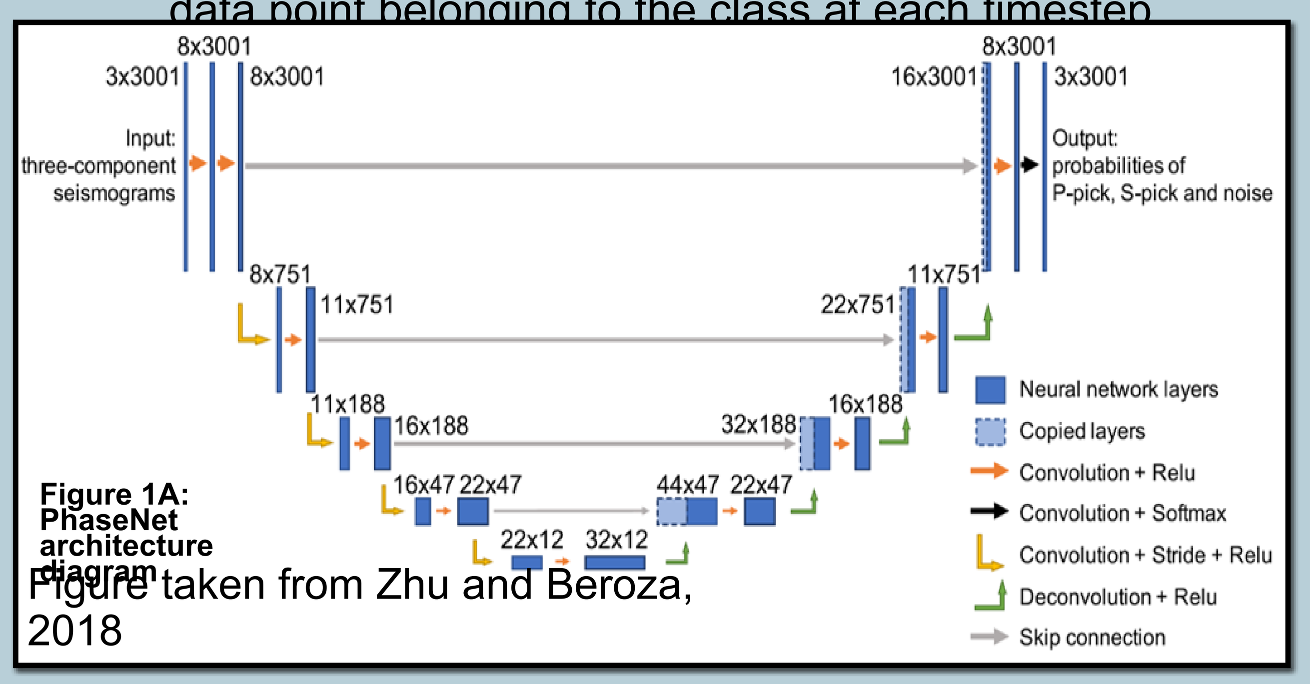


Figure 1A: PhaseNet architecture diagram

Figure taken from Zhu and Beroza, 2018

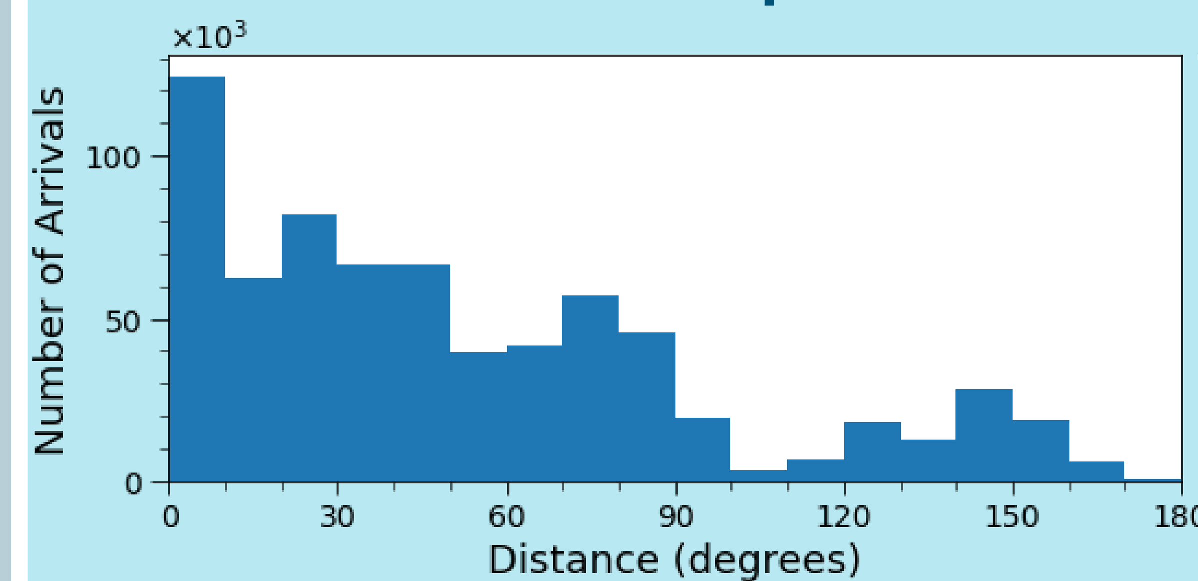
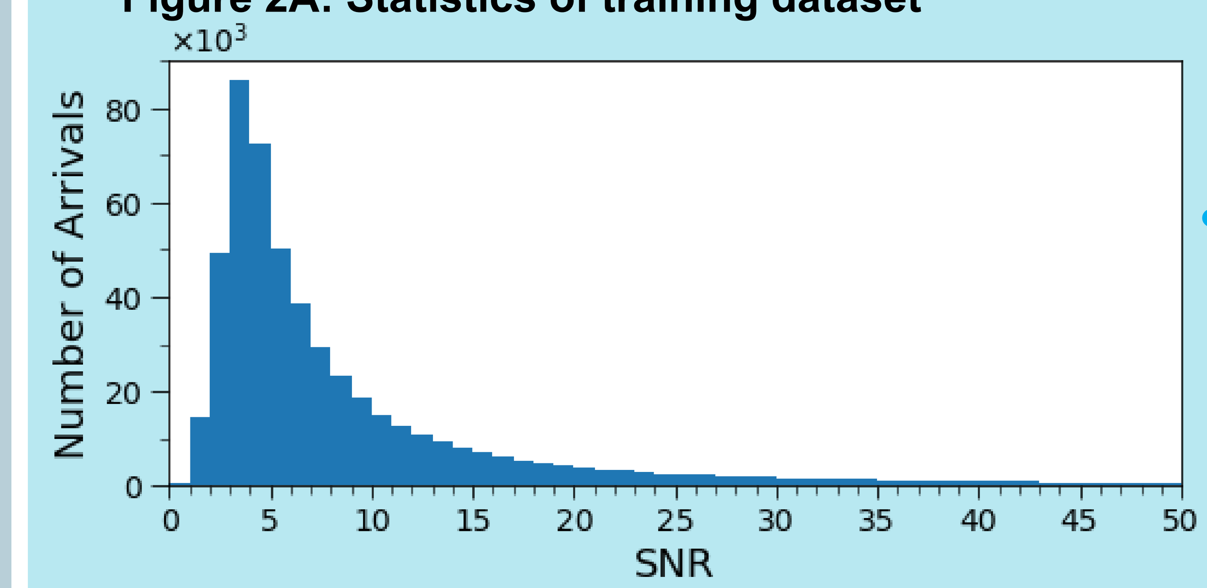


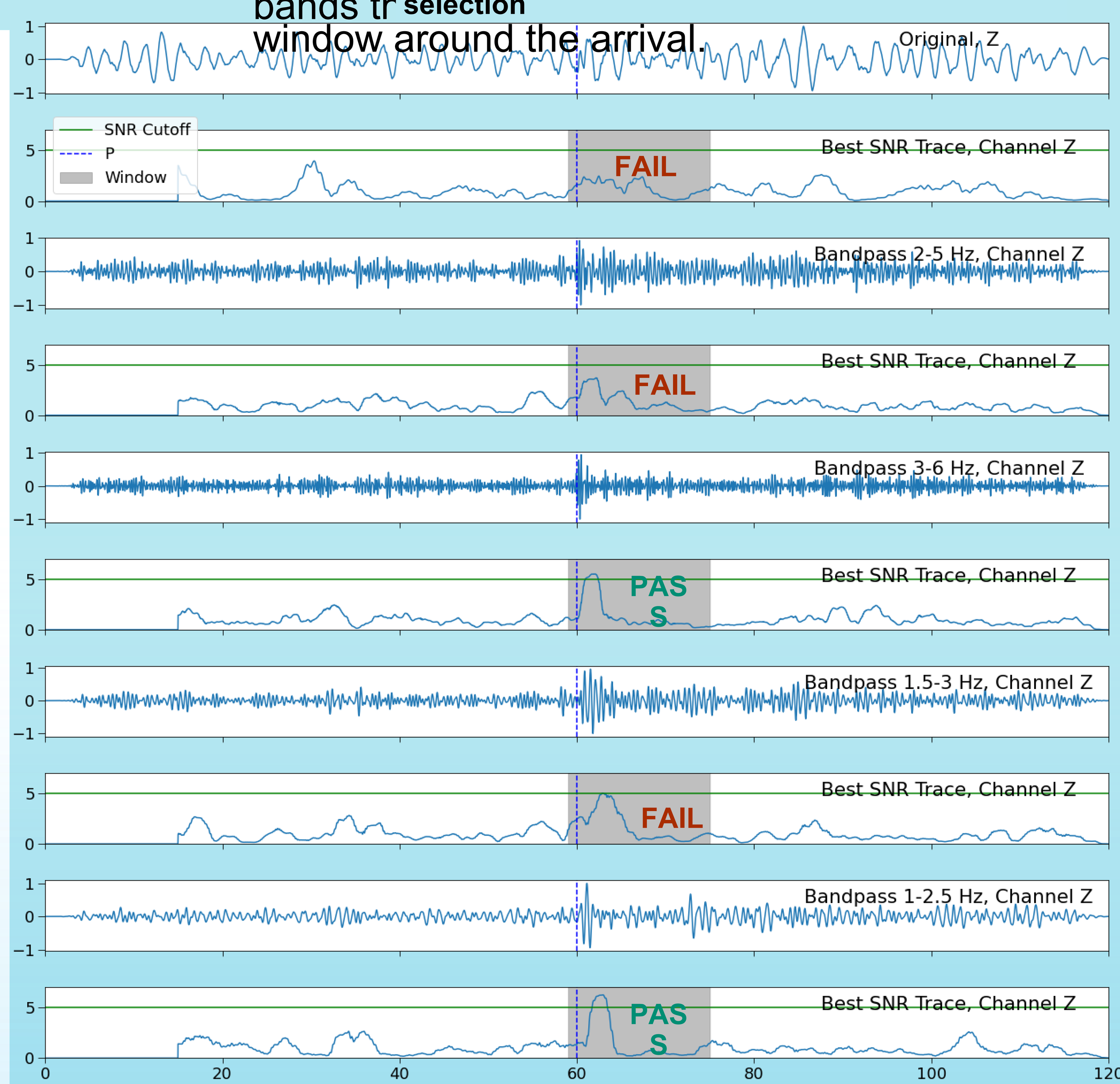
Figure 2A: Statistics of training dataset

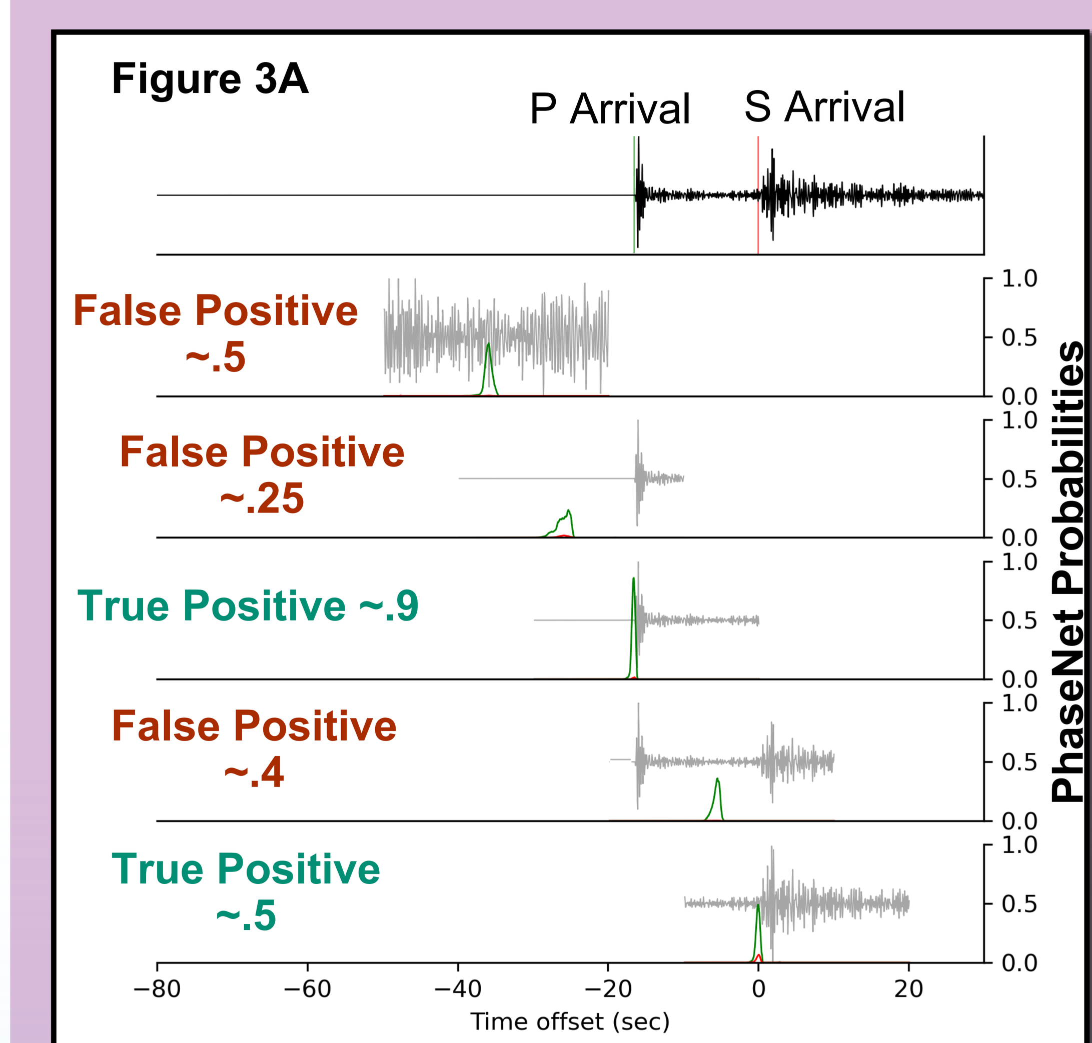


Data Details 3

- 700,681 waveforms, with simplified labels
  - All phases mapped to P or S
- Three-component station data from the from the International Monitoring System (IMS) global network. Key challenges:
  - Need a model to characterize multiple distance regimes, affecting waveform characteristics
  - Dataset is auto-curated by thresholding waveforms by their estimated SNR rather than being hand-curated.
- Quality control: Multi-filter waveform selection
  - Apply multiple bandpass filters to waveforms (mimics the International Data Centre's signal detection processing and analyst procedures)
  - Identify the best bands that have the highest SNR in a window around the arrival

Figure 2B: Example of multi-filter waveform selection



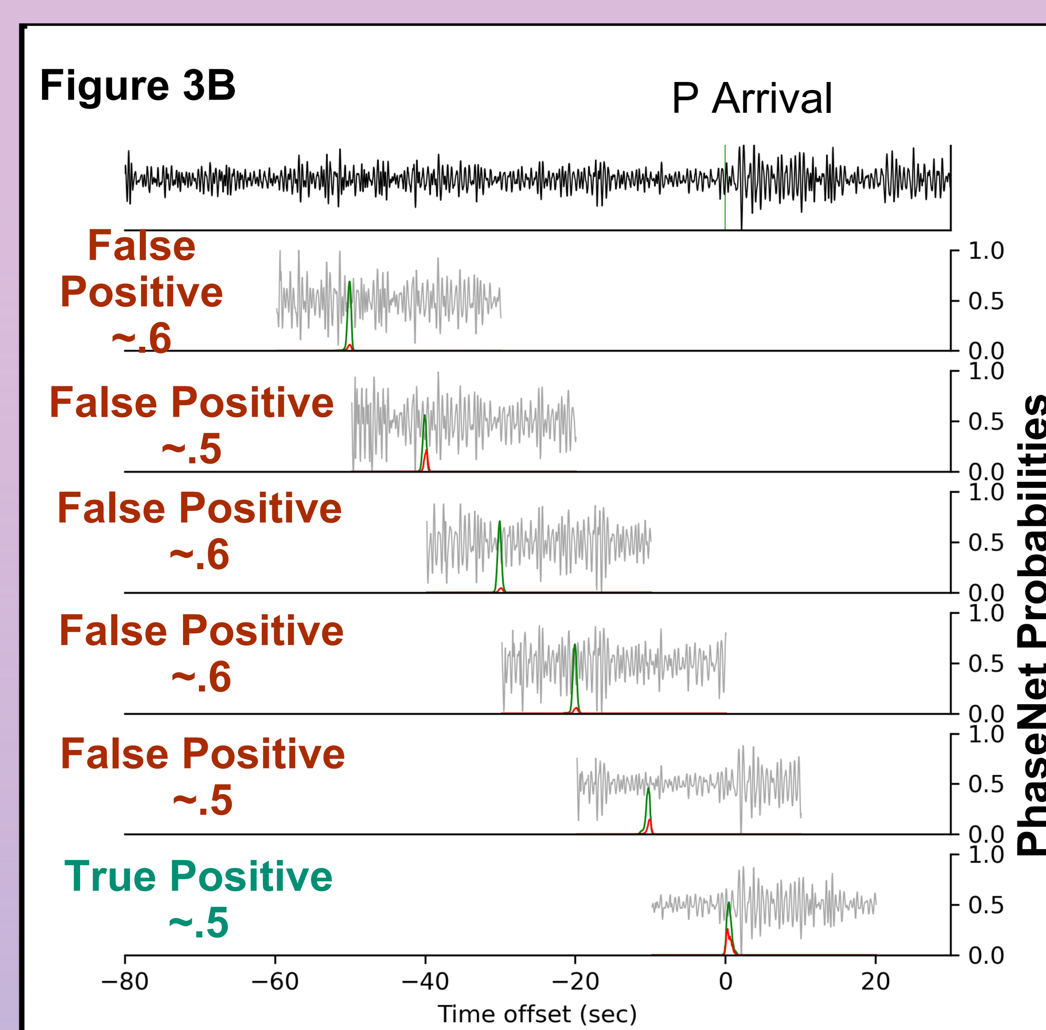


- Green traces are P probabilities
- Red traces are S probabilities

Notice the probabilities for high SNR arrivals in Figure 3A versus the probabilities for false positives in Figure 3B

Predicting on streaming data: 4

- Original model built to analyze waveform snapshots
  - Need to adapt to the case of streaming data, in which new data points are being continuously acquired
- Simulate the arrival of new data for the model to process
  - Shift the 30 second window of input data every 10 seconds along a streaming seismic channel
  - Overlapping PhaseNet output probabilities are merged by taking the point-by-point max
- Current model predicts P and S well on true signals, but has an high false-positive rate due to the high PhaseNet probabilities on data windows with no signal (Figure 3B.)



Results and Conclusion 5

- Our primary concern with the current results is the consistent presence of high probability false positives, relative to the probability peaks for true positive signals. PhaseNet is not adequately recognizing the context in which it is making these mistakes.
- Potential solutions to test:
  - Changing normalization schemes
  - Larger input window for training
  - Larger convolutional kernels for more context
  - Revise our input window scheme (e.g. CRED [Mousavi et al., 2019])

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

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.
SAND No. ________

Sandia National Laboratories