Crescent Project Report - CRESCENT SEED GRANT PROGRAM 2023-2024

A deep denoising algorithm targeting oceanic noise in seafloor seismometer data

Joseph Byrnes, Northern Arizona University.

I proposed to improve the recovery of signals of interest recorded in ocean bottom seismic data with machine learning. Subduction zones represent a major hazard to large population centers, but typically lie offshore. Seismic data collected offshore are both intensive to collect and feature unique noise sources that can obscure the signals of interest. Computational approaches to maximizing return require no additional cost to deployments and can be applied to existing datasets. This project tailored neural networks to different types of ocean bottom seismic data. The original proposal was to develop denoising techniques and to evaluate if the denoised data are reliable for routine measurements on local earthquakes. The algorithms are overall successful at their assigned task, but networks that make direct measurements are preferred over generic denoising techniques for some critical applications.

• Did you accomplish what you set out to with this project?

Yes (mostly) – but the proposed technique shows limitations.

The foundation of this project was to develop denoising algorithms for local earthquakes (that is, earthquakes within approximately 100 km of the recording station), to evaluate if denoising is useful or misleading, and then to explore how well techniques generalizes to other types of data such as tremor or teleseismic earthquakes. Supplemental funds were provided at NAU to extend the project to active source seismic data recorded on OBSs using the same workflow. The network I proposed to use was the short-time fourier transform (STFT) based approach described in *Thomas et al.*, 2023 along with noise sampled in windows defined by *Janiszewski et al.*, 2023, and to explore the potential for using the pressure channels to enhance machine-learning applications above what can be done on land. A purely time-domain approach was ultimately included as well (*Yin et al.*, 2022), along with a novel design that operates on the discrete wavelet transform (DWT). All three approaches are natively multi-channel.

Two datasets were used for training the two branches of the project (following the supplemental funds for active source data). For local earthquakes, the PNW earthquake dataset (*Ni et al.,* 2023) provided approximately 30k high signal-to-noise ratio signals for earthquaks. Additionally, about 100k samples of noise from broad-band OBSs deployed in Cascadia were collected from windows of time provided by H. Janiszewski. For active source data, the training set had to be built from scratch. 350k high signal-to-noise ratio arrivals from the R/V Langseth were collected by an automated approach along with 440k samples of OBS noise across 18 active-source experiments.

A problem arose here in which datasets of earthquakes recorded on OBSs for training AI networks have been collected and published, but almost none of these collected earthquakes have high signal-to-noise ratios below about 2.5 Hz on all components but especially on the pressure channel, precluding the use of the pressure channel for training denoisers (differential pressure gauges have rapidly decreasing sensitivity above 2.5 Hz and hence often show no signal, even when the seismic data are excellent). Pressure data was fully included for active-source data, as the instruments in those experiments use a hydrophone instead of a differential pressure gauge and the signal is concentrated above 3 Hz.

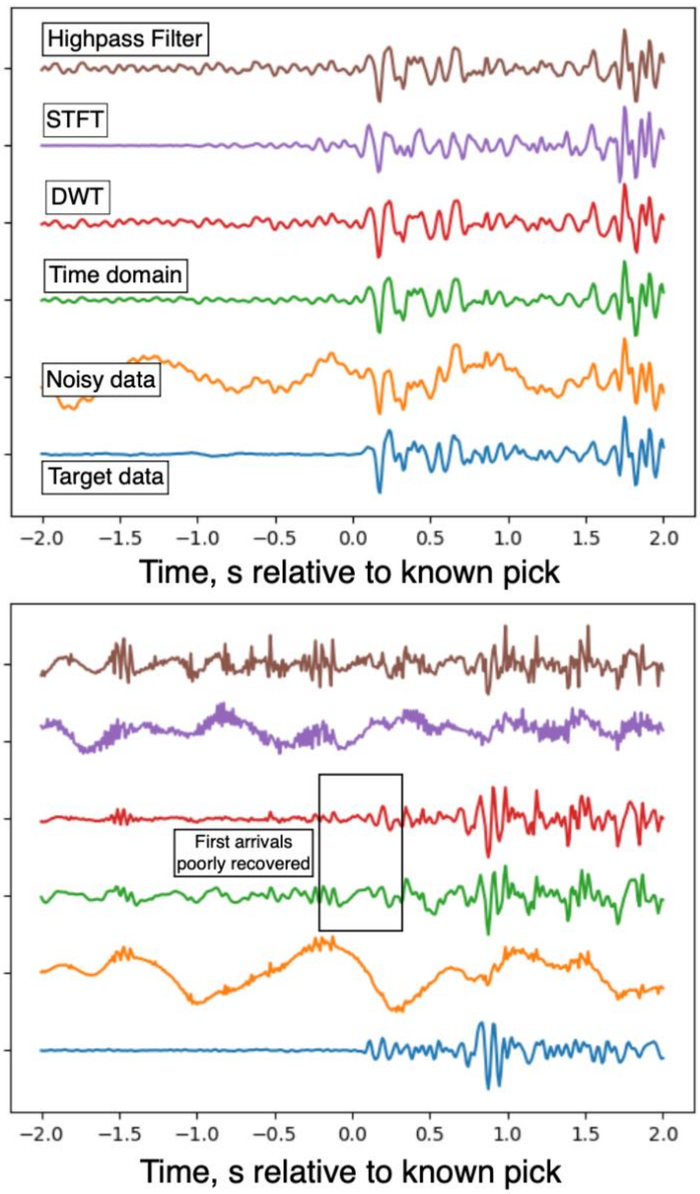

Figure 1. Denoising examples. Example in the top panel is easy to denoise, but barely outperform a simple filter. In the bottom filter, two networks partly recover the waveform, but distort the first arrival.

I found that denoising networks can be successfully trained and can perform well for waveform reconstruction - but can be unreliable. **Denoising networks reconstructs the bulk waveform, but the most important part is the first arrival** and this small portion can contribute very little the total L2 misfit that the networks aim to minimize. Two examples are shown in Figure 1. For a mild denoising example, time-domain, wavelet, and STFT methods correctly reconstruct the waveform but the performance is only marginally better than simple bandpass filtering, as the ocean noise is mostly long period. A more difficult denoising target, in the bottom panel of Figure 1, shows that the time-domain and wavelet networks greatly outperform a simple filter but do not necessarily recover the important first arrival as well as the coda. In general, the STFT denoiser performed poorly but this is likely due to this network having far more parameters than the other two, and the PNW training set, while large, may not have been large enough for this network style.

The same issue arises with first arrivals in active source data, as shown in Figure 2. This is an extreme example – the vast majority of denoised waveforms for the active source data are excellent. The RMS of the misfit to denoised picks to the true picks are only 86 ms (at the same signal-to-noise ratios used for training), while the example in Figure 2 is off by 550 ms. However, this set of misfits has a kurtosis of 26 (a normal distribution has a kurtosis of 3), indicating a very heavy-tailed distribution driven by outliers as in Figure 2.

- If not, what did you do differently and how did you account for the changes?

Direct picking networks are more reliable than denoising networks for certain critical parameters.

A network for direct measurements with errors was trained in parallel with the denoising networks. The training sets contain the information of interest – arrival times for the active source data, and arrival times and first-motion polarities are cataloged along with waveform in the PNW dataset; amplitudes were calculated on the fly. The direct measurement network is the same as the time-domain network, except with the final deconvolution layers replaced by a transformer that pools information to produce discrete information (along with uncertainties, calibrated by using negative log likelihood as the loss function for picks and amplitudes and binary cross entropy for polarities). A wavelet approach was tested but failed to detect first-motion polarities.

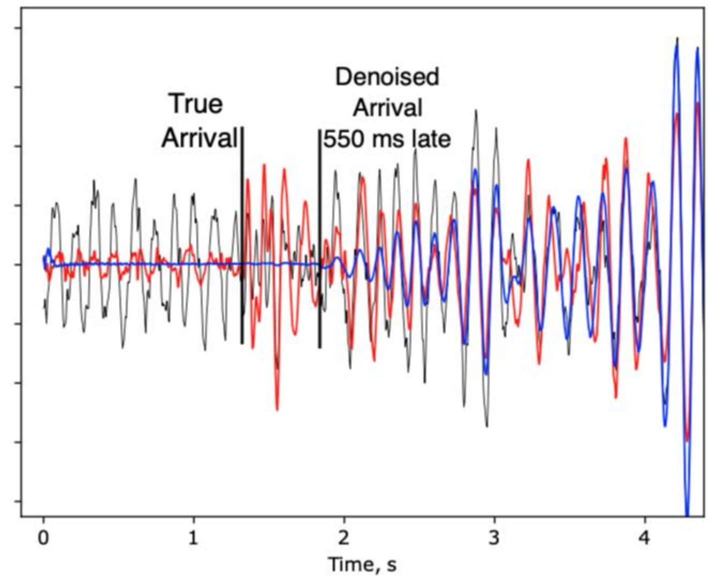

Figure 2. An example of misleading denoising for active source data. Red is a clean arrival of a shot from the R/V Langseth, contaminated with noise in black. Blue is a denoised trace.

Figure 3 compares all the metrics of interest for local earthquakes and compares the performance of simple filtering, neural networks that operate on data on time-domain, DWT, and STFT data, along with measurements from the direct picking network. Waveform reconstruction is shown as explained variance (EV), with 1 implying perfect reconstruction and 0 meaning no correlation between the target and denoised traces. **Neural networks work well at waveform reconstruction and outperform a bandpass filter**, though note that their relative performance may represent the relative size of the networks – the larger STFT based network also has worse validation loss relative to training loss and would likely work better if the training set were expanded. Three quantities at the first arrival are evaluated in the following three panels – arrival times, (log) amplitudes, and the polarity of the first arrival. All denoising techniques are shown to sometimes distort the first arrivals and **direct picking on the noisy waveforms always outperforms denoising**. This plot is technically accurate but likely overestimates the performance of the denoisers – the bottom panel in Figure 1, for example, technically features an accurate first motion polarity but would be hard for an analyst to identify independently. For some applications, the waveform is what is needed – but for general earthquake analysis, the direct approach is likely preferrable.

For active source data, a **direct picker solves the outlier problem found with the active source denoiser** – RMS misfit to the test set is 50 ms and the kurtosis is reduced to only 6 (a normal distribution has a kurtosis of 3). A secondary network was trained on top of this first network to pick relative arrival between traces, and a system of linear equation solves for a combined arrival time. An application is shown in Figure 4.

• What is the next step for development of this project/ priority?

The text of the original proposal was focused on application to local earthquakes, and then generalization would be explored. **Generalization across classes of seismic signals is terrible** – teleseismic events are not recognized, and the active source networks cannot identify earthquakes nor

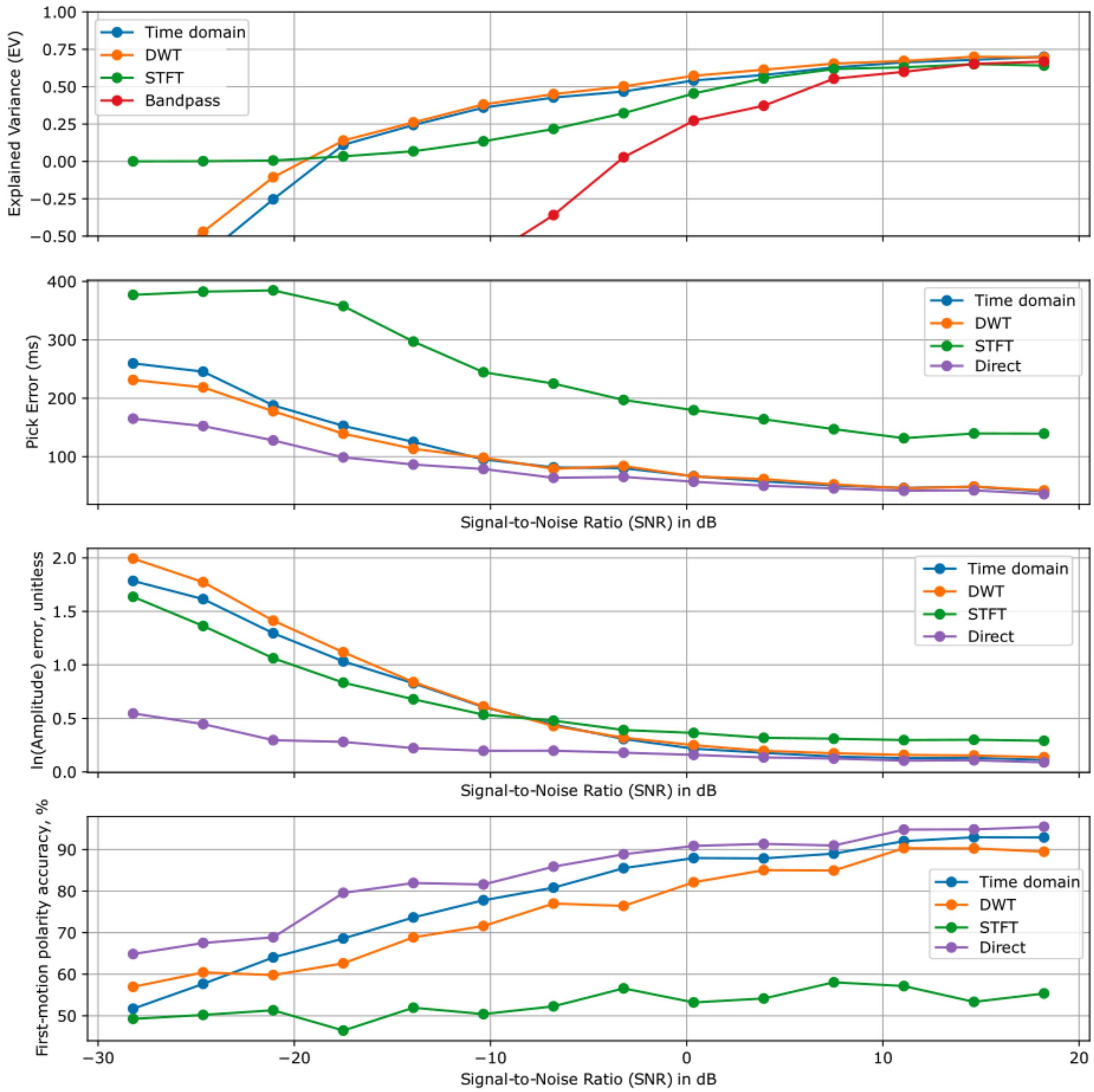

Figure 3. Comparison of different approaches along 4 metrics. From top to bottom, metrics shown are the explained variance, the accuracy of the first arrival, the log of amplitude of the *P* wave, and the polarity of the first motion

vice versa. The proposal specifically mentioned attempting to detect non-volcanic tremor and based on these results, the networks are currently being adapted to detect non-volcanic tremor on OBSs by a master's student at UT Dallas, where the PI is starting a new position in the Fall. Similar efforts will also be adapted to teleseismic data, for which collecting the training sets and any potentially good data on the pressure needs will need to be explored – semi-supervised learning and/or generative networks are both potential avenues for pushing machine-learning techniques over this hump. Application of the active source pickers is also underway, with the CASIE dataset from Cascadia a prime target.

• Where did you publish/ present on this work?

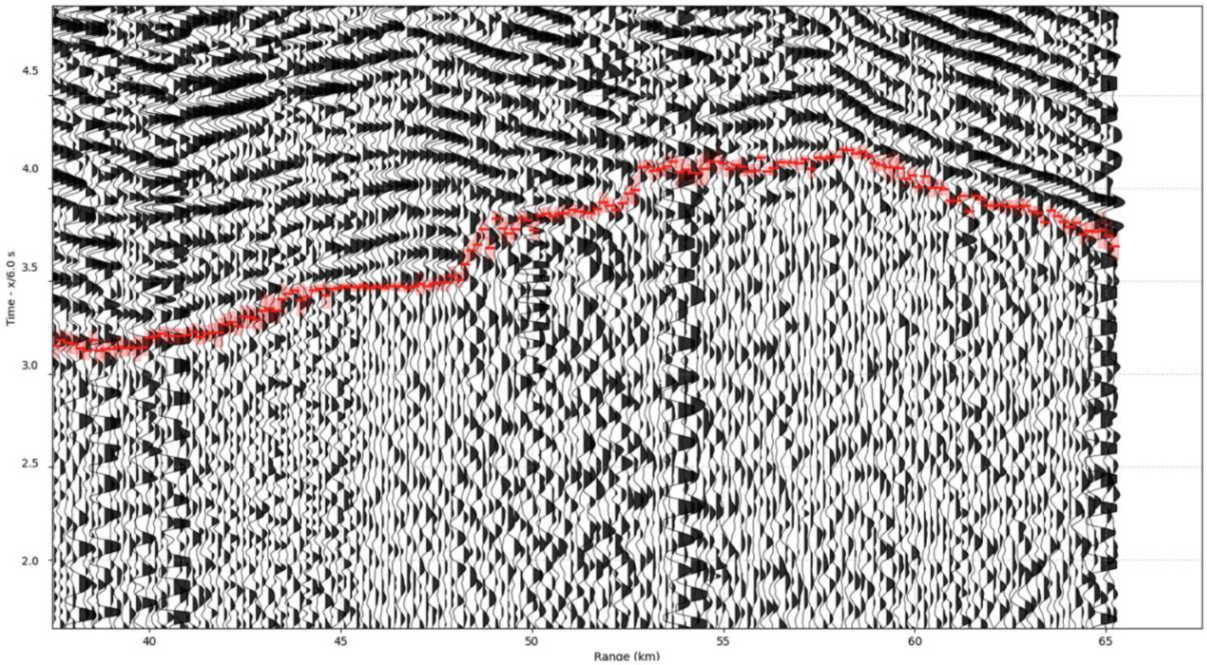

Figure 4. Example of picking active source data with relative arrival times included. Data are vertical component seismograms from the Santorini experiment recorded on OBS 103, Line 5. Red bars are machine-learning picks, and shaded regions are the predicted 1-sigma errors.

Both the local earthquake and active source networks described above are in prep for a target of 2025 submission. Results are mature, as detailed above, but some work remains to finalize into publications. For publication all four networks local earthquake networks will be training on a much larger training set and reevaluated. The results have been presented in four seminars by the PI, 3 at universities and at the SSA-funded Arizona Collaborative Consortium for Earth and Space Science.

References

Janiszewski, H. A., Eilon, Z., Russell, J. B., Brunsvik, B., Gaherty, J. B., Mosher, S. G., Hawley, W. B., Coats, S. Broad-band ocean bottom seismometer noise properties, Geophysical Journal International, Volume 233, Issue 1, April 2023, Pages 297–315, https://doi.org/10.1093/gji/ggac450

Ni, Y., Hutko, A., Skene, F., Denolle, M., Malone, S., Bodin, P., Hartog, R., & Wright, A. (2023). Curated Pacific Northwest AI-ready Seismic Dataset. Seismica, 2(1). https://doi.org/10.26443/seismica.v2i1.368

Thomas, A., Melgar, D., Dybing, S. N., & Searcy, J. R. (2023). Deep learning for denoising High-Rate Global Navigation Satellite System data. Seismica, 2(1). https://doi.org/10.26443/seismica.v2i1.240

Yin, J., Denolle, M. A., He, B. A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms, Geophysical Journal International, Volume 231, Issue 3, December 2022, Pages 1806–1822, https://doi.org/10.1093/gji/ggac290