Neural Implicit Compact Representation to Compress Distributed Acoustic Sensing Data

Yiyu Ni[1*], Shaowu Pan[2], J. Nathan Kutz[2], Bradley P. Lipovsky[1], Marine A. Denolle[1]

[1] Department of Earth and Space Sciences, University of Washington, Seattle, WA, 98195

[2] Department of Applied Mathematics, University of Washington, Seattle, WA, 98195

*niyiyu@uw.edu

SCEC AN NSF+USGS CENTER

Background

- The ***Distributed Acoustic Sensing (DAS)*** is a new seismic observation method that utilizes repeated laser pulses and measures changes in the phase of backscattered light to measure the strain rate along with an optic fiber.
- Parameterized by neural networks, **Implicit Neural Representation (INR)** is a machine learning technique that represents data in a compact space. An INR model is defined and trained with data sampled from the original representation, and the data is then **represented by the parameters** within the INR model. In order to reconstruct the original representation, parameters like the row index of the matrix serve as the input to query the data from the INR compact representation. INR works as a lossy data compression and transmission method, where data can be reconstructed with loss.

OOI DAS

OOI DAS data was collected during the maintenance of the **OOI Regional Cabled Array** off central Oregon, November 1-5, 2021. The cables were connected to two **Optasense QuantX DAS Interrogators**, and 9.7/5.2 TB data raw data were recorded on the north and south cables, respectively.

For the south cable, there are totally 47500 channels recording at 200 Hz along the cable spacing by 2.07 meters. Assuming a double precision recording (8 bytes), OOI DAS generates data at a rate of ~72 MB/s, which makes real-time transmission impossible.

Figure 3 shows a snippet of OOI DAS data from the south cable segment highlighted by a yellow rectangular in Figure 2. It shows data of three cable segments for 1 minute in its original representation.

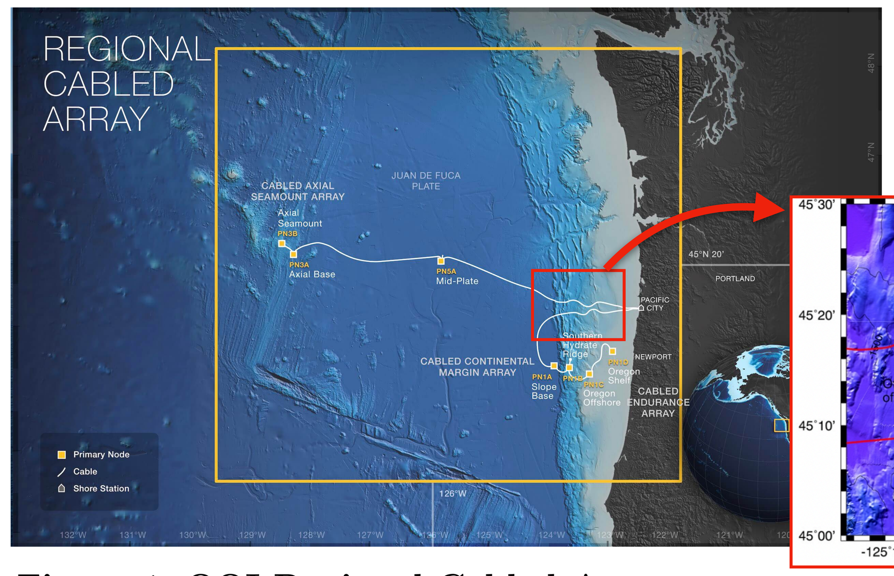


Figure 1. OOI Regional Cabled Array

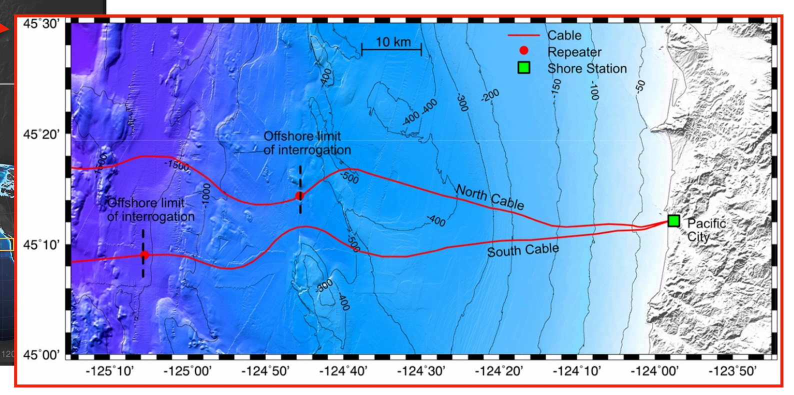


Figure 2. OOI DAS Cable Geometry

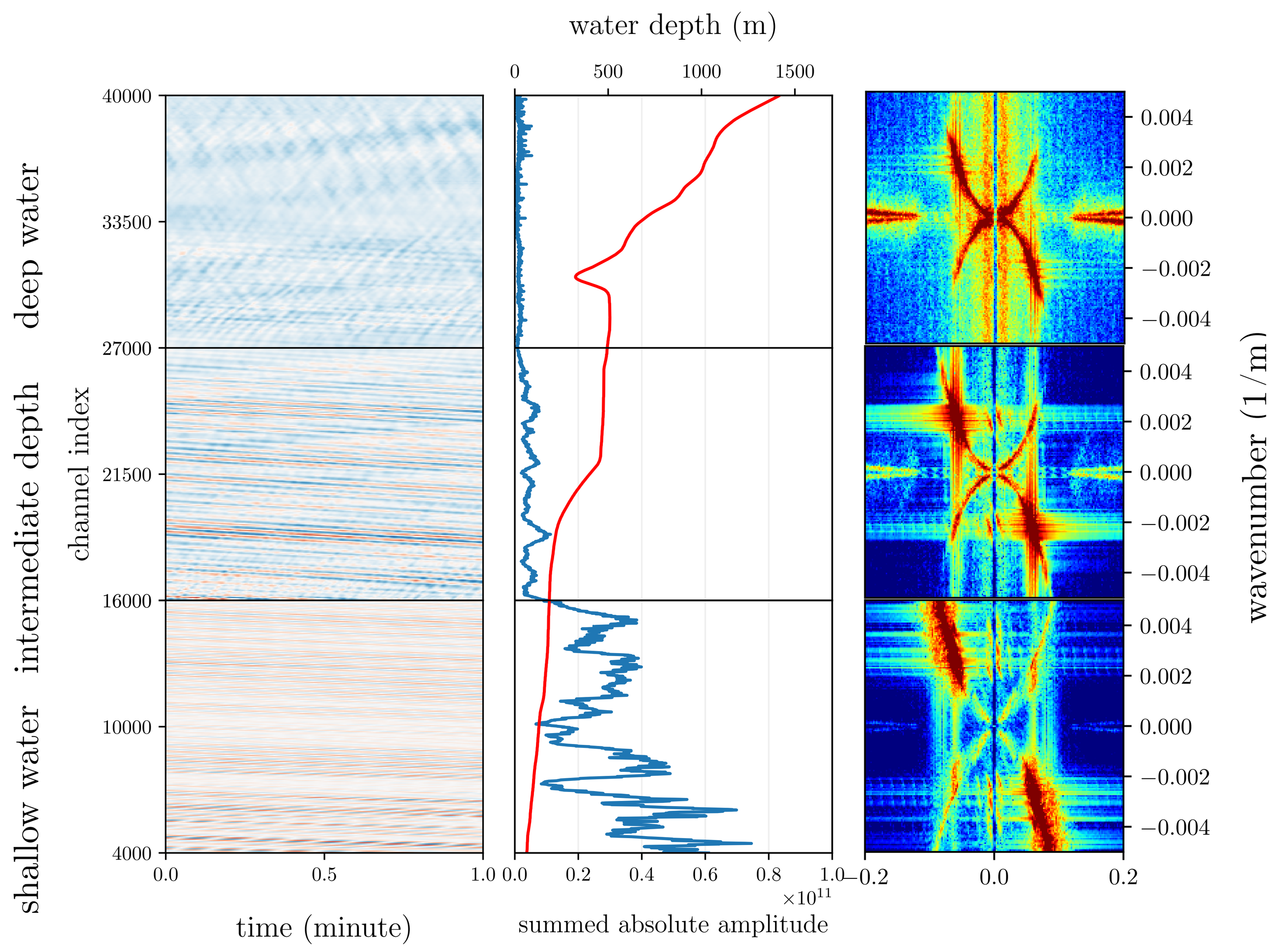


Figure 3. Left column: One minute of OOI DAS data in its original representation. Data demeaned by channel. Middle: summed absolute amplitude of each channel. Right: data in frequency-wavenumber domain.

Method

In order to showcase this method, we chose a 15-minute recording from OOI DAS data. Since different segment of the cable have different dominant frequency and dispersion relation (Figure 3 and 4), we split the whole cable into three segments: index 3000-15000, 15000-25000, and 25000-40000. We train three models independently for each cable segment.

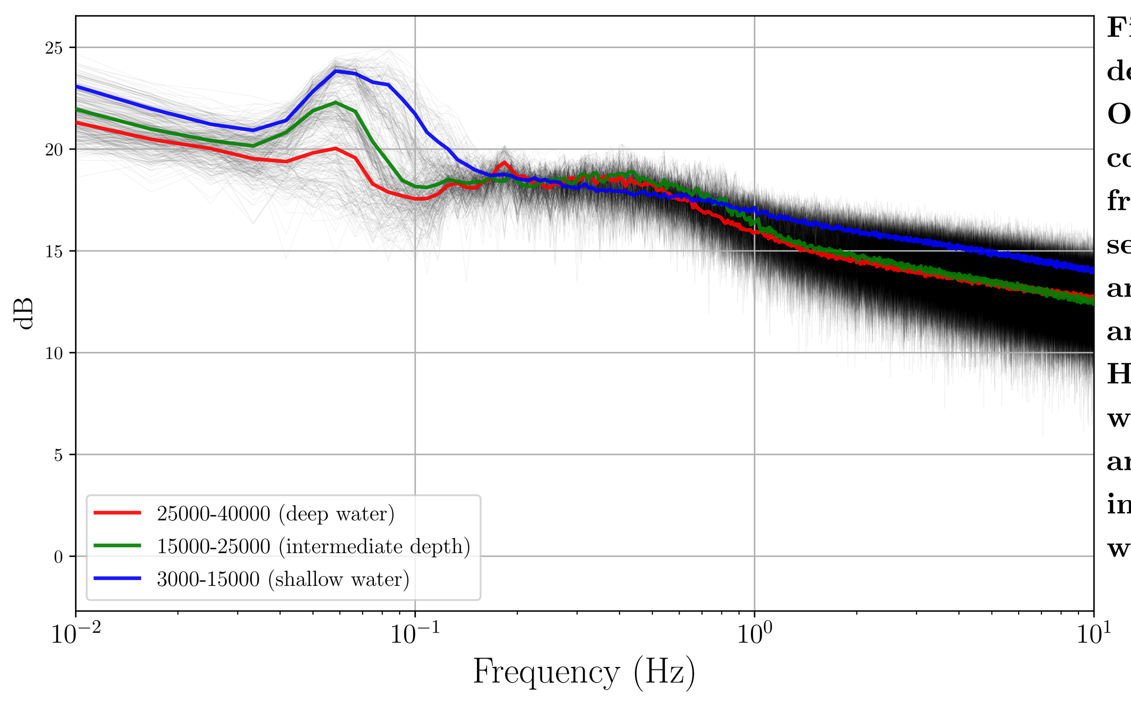


Figure 4. Power spectrum density of two minutes OOI data after applying a cosine taper. Dominant frequency of all cable segments is around 0.6 Hz, and deep water has another peak around 0.19 Hz. Data from shallow water has higher spectrum amplitude than intermediate and deep water cable.

Figure 5 shows architectures of ***SIREN[1]*** and ***Random Fourier Feature Network[2]***, which are the INR models used in this research. We created the training dataset by selecting data points from the original representation data matrix. In this 15-minute data compression test, there are 8.6, 7.2 and 10.8 million data points from shallow, intermediate and deep water segments, respectively. We define and train a SIREN model with 0.33 million parameters on a single A100 GPU for 30 epochs using Adam optimizer and MSE loss function.. The same dataset is used to train a random Fourier feature network with 0.39 million parameter. Instead of having a traditional train-validate-test process to avoid overfitting and underfitting, we train and overfit both models using one training dataset.

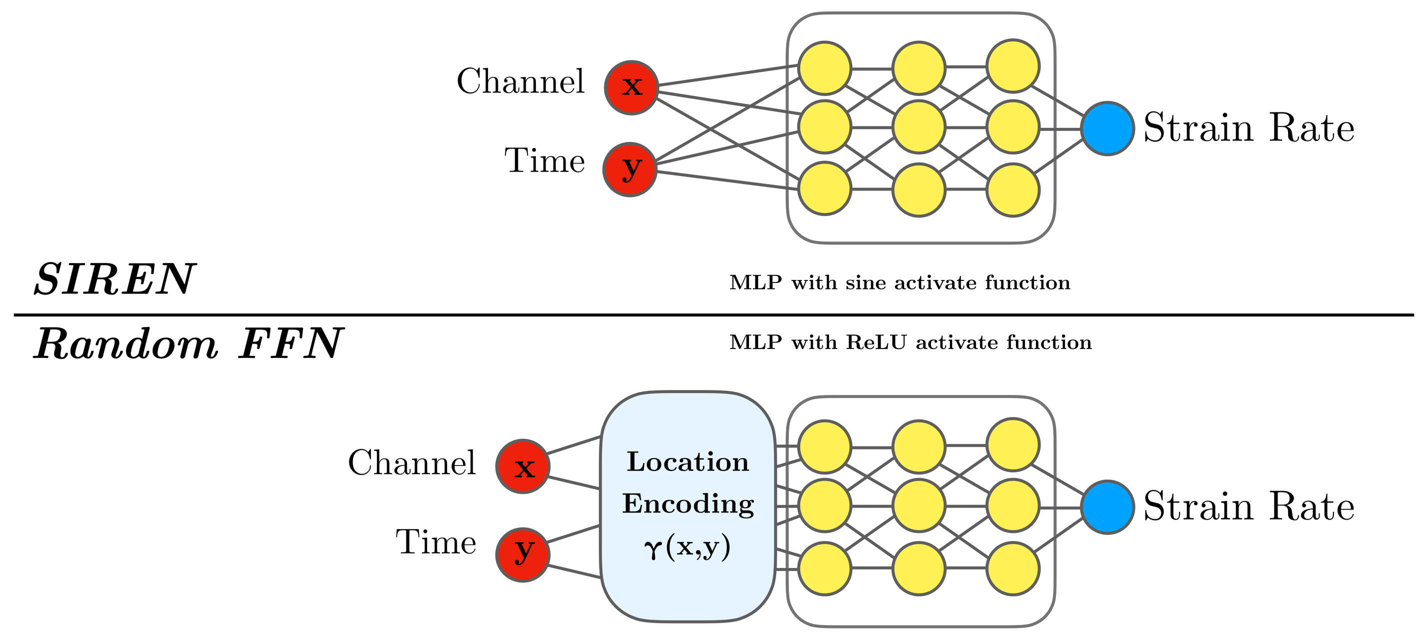
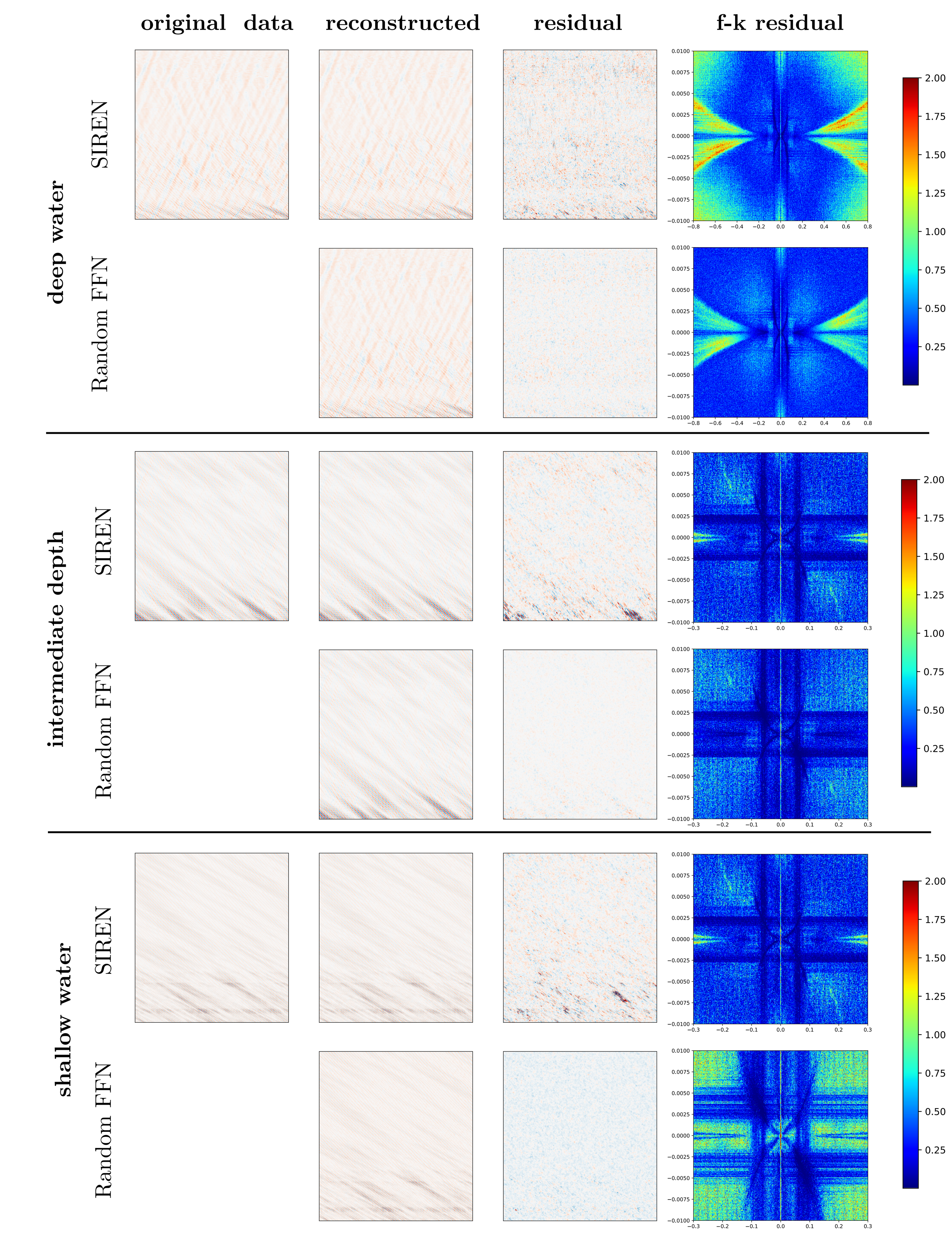


Figure 5. Architecture of SIREN and Random Fourier Feature Network.

The overfitted model compactly stores the data from their original representation. The INR model represents the data with trainable weights, getting ~3% compression ratio for this test. In order to reconstruct the complete but lossy data, the channel and time index serve as the input to query the data from the INR model.

Result



Discussion

- The signal from the ***Ocean Surface Gravity Wave[3]*** is well-recovered in lower-frequency, and the high-frequency signal suffers more loss.
- SIREN works well in both shallow and intermediate water depth. But since SIREN favors a continuous representation both in space and time, getting worse reconstruction in deep water where continuities across channels are poor.
- Random FFN is computationally expensive than SIREN, but works better in reconstructing higher frequency signal, e.g., in deep water.

Ongoing work

- Test the data fidelity of transient and high frequency (earthquake and whale) and ambient signal.
- Test this method on urban DAS experiment.

Reference

1. *Sitzmann, V., Martel, J., Bergman, A., Lindell, D., & Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. Advances in Neural Information Processing Systems, 33, 7462-7473.*
2. *Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., ... & Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems, 33, 7537-7547.*
3. *Williams, E. F., Zhan, Z., Martins, H. F., Fernández-Ruiz, M. R., Martín-López, S., González-Herráez, M., & Callies, J. (2022). Surface Gravity Wave Interferometry and Ocean Current Monitoring With Ocean-Bottom DAS. Journal of Geophysical Research: Oceans, 127(5).*