Implicit Neural Compact Representation of the OOI DAS Data

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

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

[2] Department of Applied Mathematics, University of Washington

*niyiyu@uw.edu

OOI REGIONAL CABLED ARRAY

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 1 shows a snippet of OOI DAS data from a south cable segment highlighted by a yellow rectangular in Figure 2. It shows data of 6000 channels for 1 hour in its original representation, which is a matrix of shape 6000 X 720000.

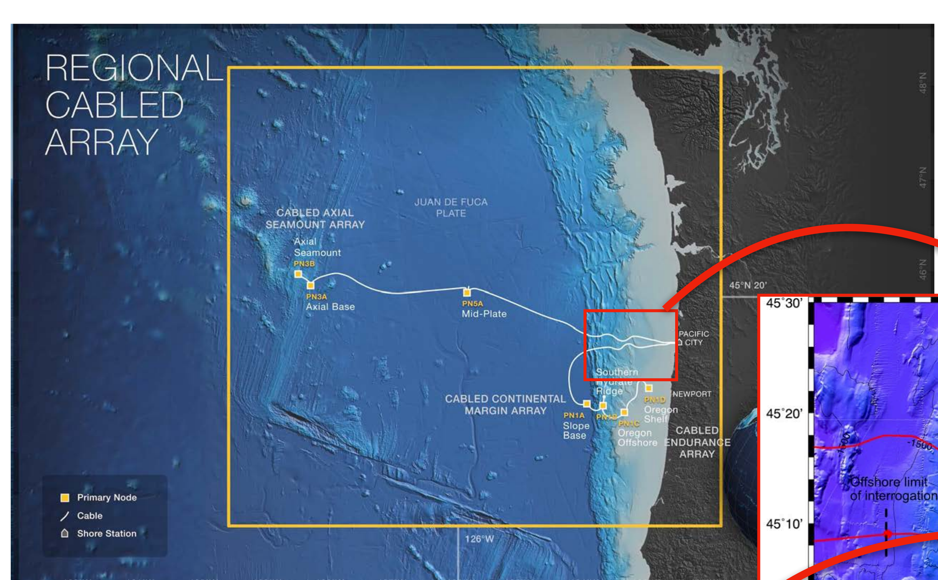

Figure 1. OOI Regional Cabled Array

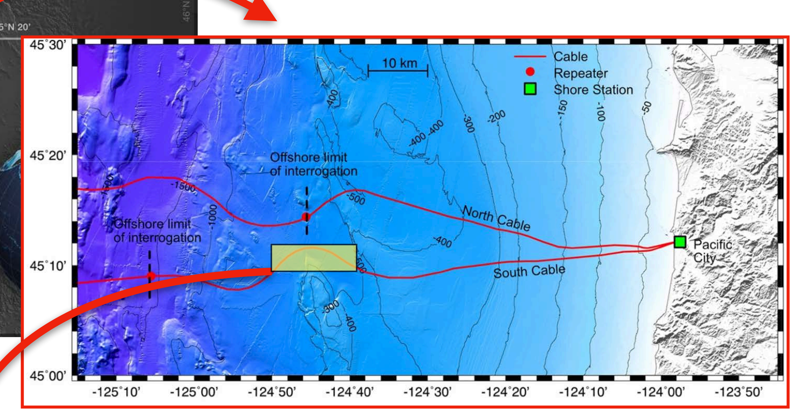

Figure 2. OOI DAS Cable Geometry

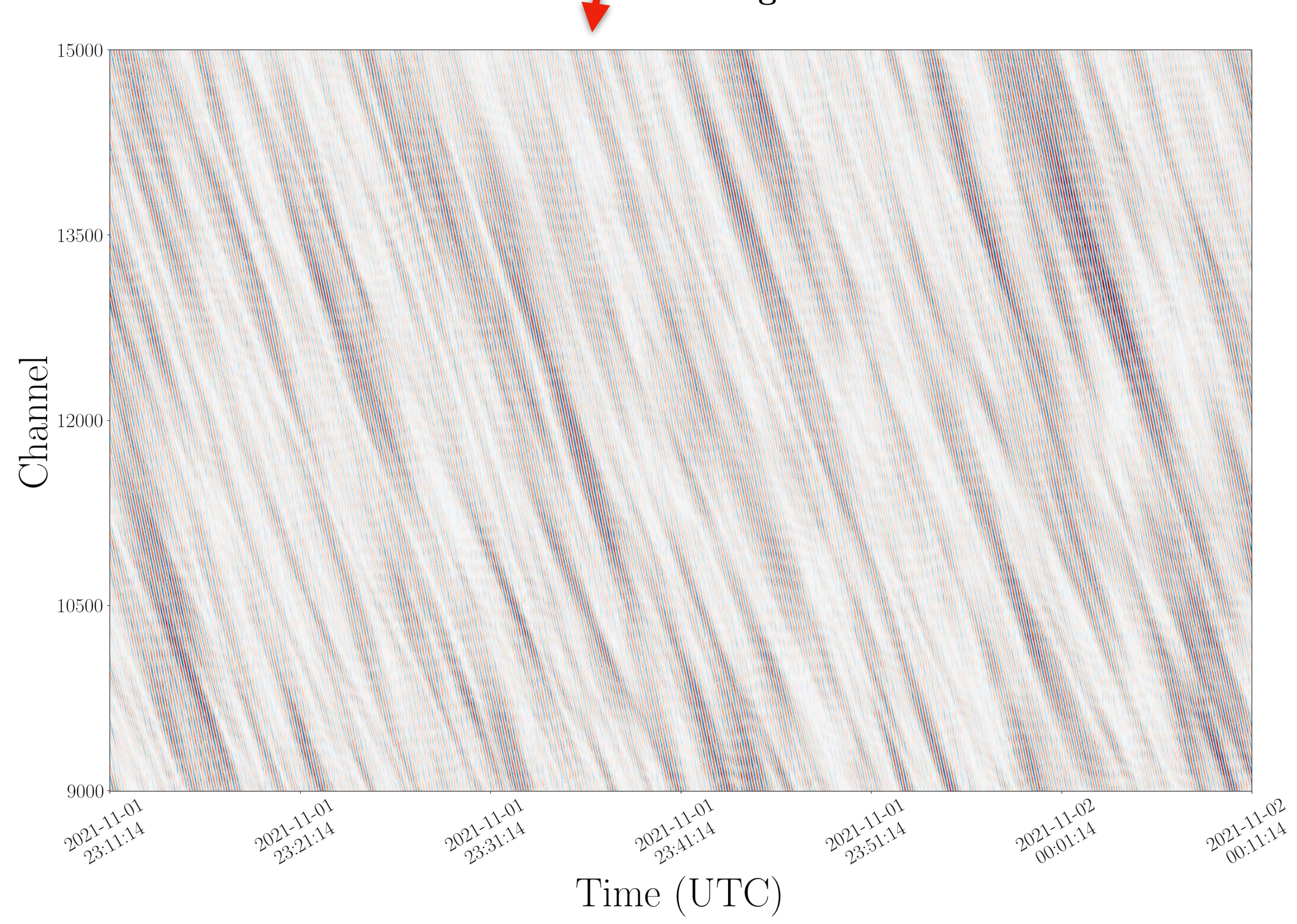

Figure 3. OOI DAS South Cable 1-hour recording demeaned by channel.

Method

To present the method, we compress a 10-minute OOI DAS data with ***SIREN***[1] model. The data in the original representation requires ~1.5 GB disk space.

1. Sampling

We randomly select data points from the original representation data matrix. There are 72 million data points per minute, and we are testing various sampling rates in this work.

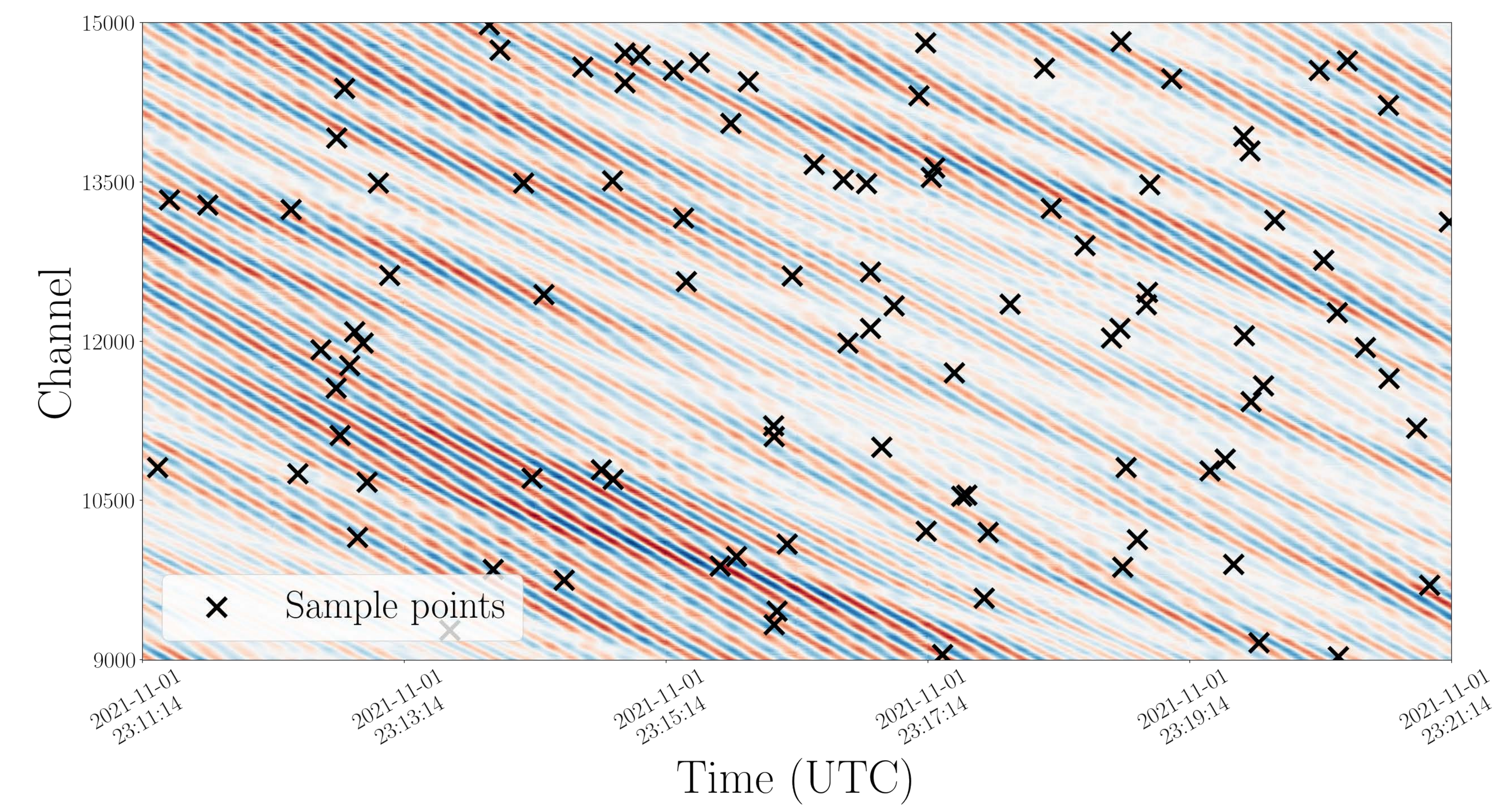

Figure 4. 10-minute OOI DAS data in original representation. Black crosses represent 100 of the sampling points.

2. Training

Using the data sampled in the first step, an INR model is trained at the site. We define and train a SIREN model with 186k parameters as the INR model. We train the model on a single RTX 3090 GPU for 20 epochs using a batch size of 200k, Adam optimizer and learning rate of 5e-5.

3. Reconstruction

As the well-trained INR model compactly stores the data from the original representation, the complete but lossy data could be reconstructed. The channel index, as well as the time serve as the input to query the data from the SIREN model. Figure 5 shows the 10-minute DAS data and its residual reconstructed from a SIREN model (trained with data sampled at 0.5 million points per minute).

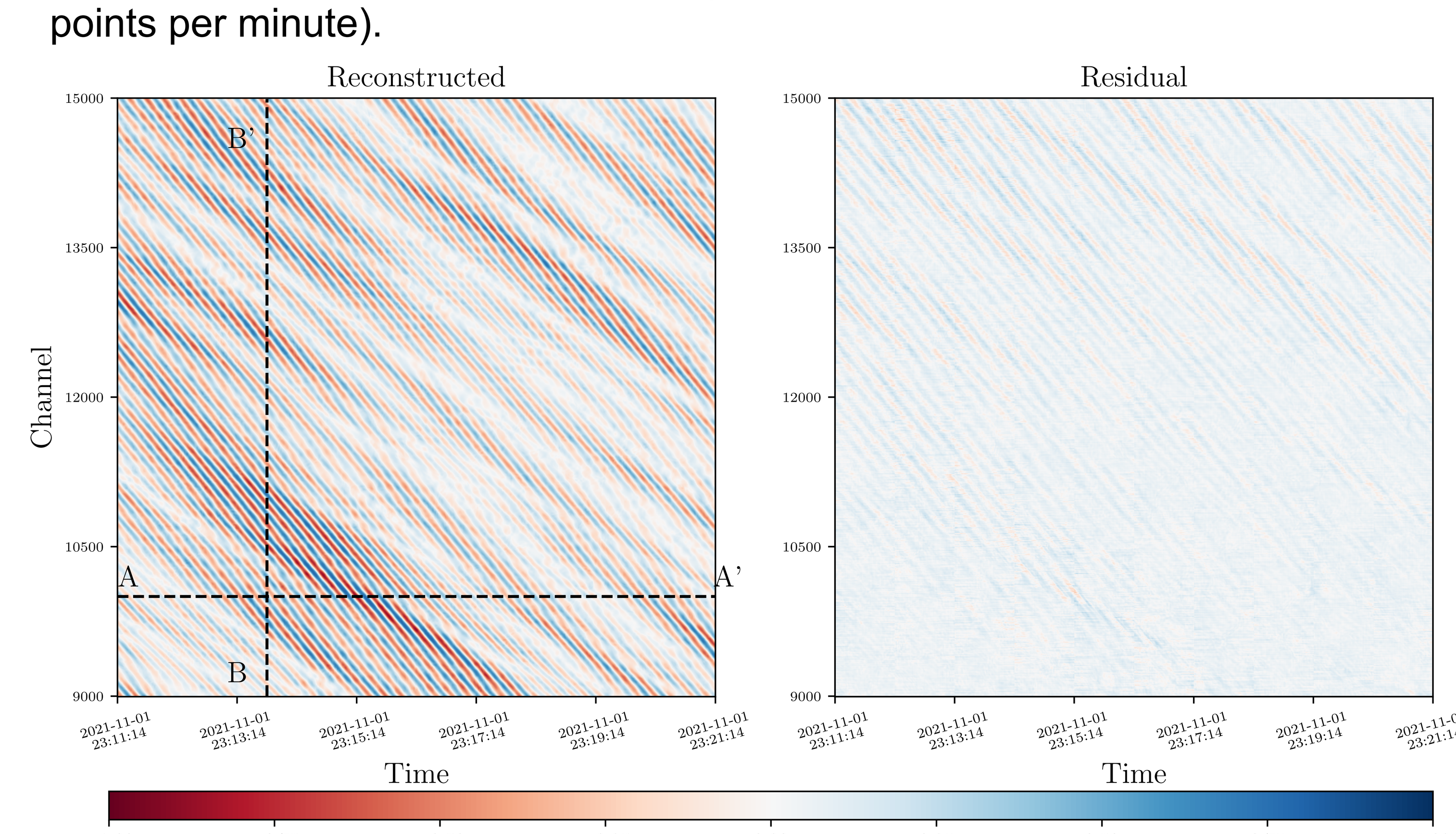

Figure 5. Full reconstruction of 10 minutes OOI DAS data snippet and its residual normalized to [-1, 1] interval.

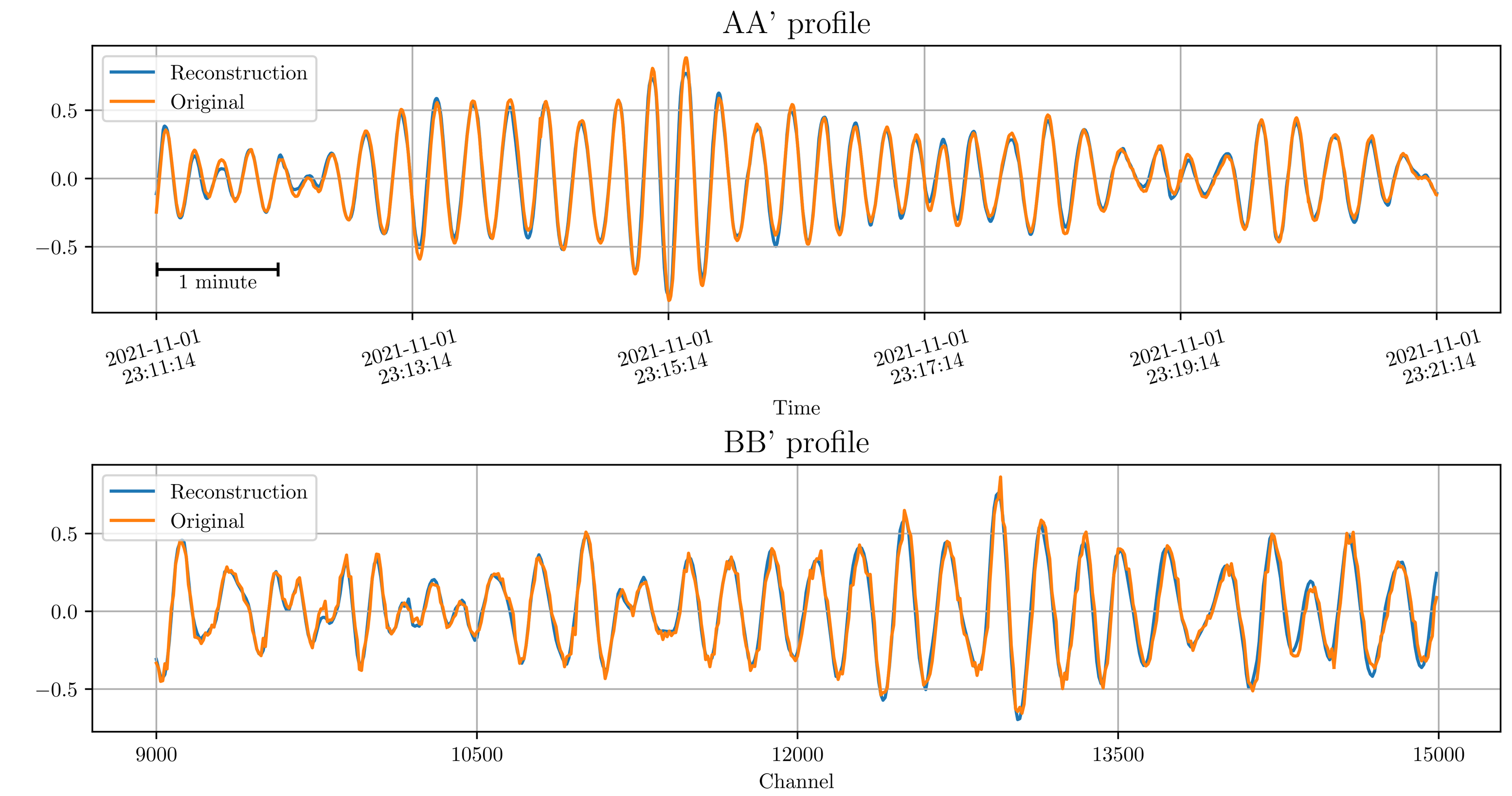

Figure 6. Comparison of original and reconstructed data on the AA' and BB' profile.

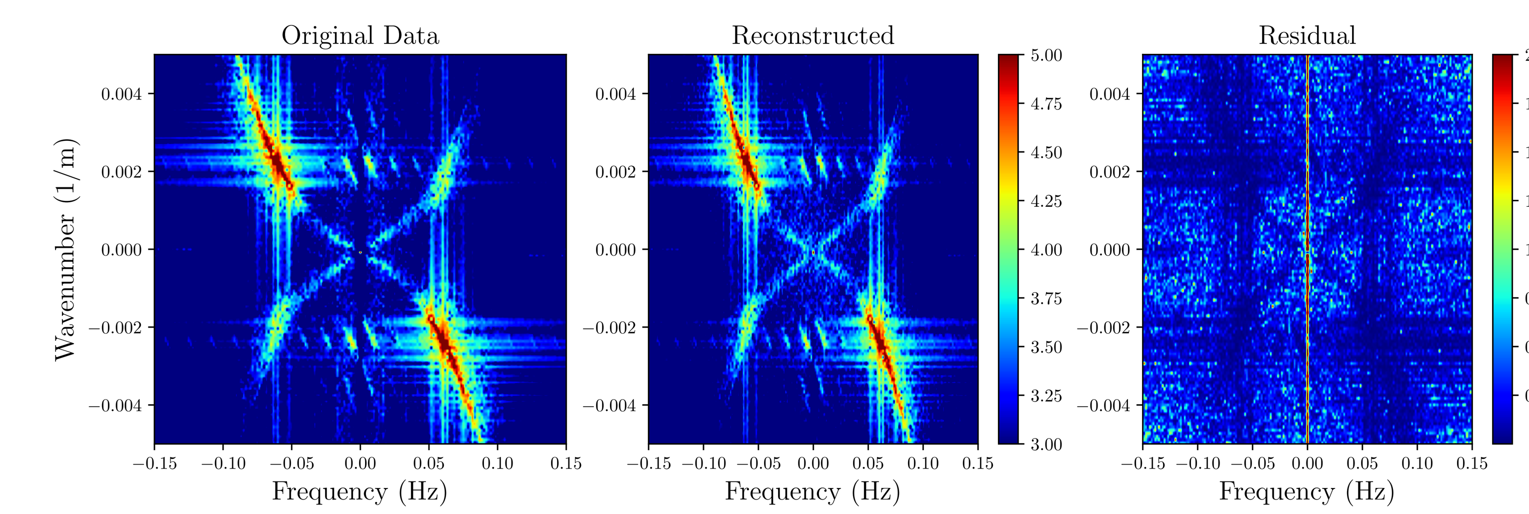

Figure 7. Comparison of original and reconstructed data in f-k domain.

Results and Discussion

- The signal from the ***Ocean Surface Gravity Wave***[2] is well-recovered in high-frequency, the low-frequency component suffocates more loss.
- Table 1 shows the results of training and reconstruction of OOI DAS data using the SIREN model with different sampling rates. As 2.2 MB disk space is required to store the model, the compression ratio is **0.142%**.

Sample rate (samples/minute)	Training set size (MB)	Sampling rate (%)	Training Time (minute)	Residual RMS
2.5 M	573	3.47	31.97	0.0521
1.5 M	344	2.08	18.50	0.0523
1.0 M	229	1.38	12.00	0.0539
0.5 M	**115**	**0.69**	**5.87**	**0.0556**
0.2 M	46	0.27	5.42	0.0579
0.1 M	23	0.13	2.67	0.0609

Table 1. Benchmark of training SIREN with various sampling rate. Tests conducted on a single RTX 3090 GPU.

Ongoing work

- Scale the INR up to data segment longer than 1-hour, 1-day, or the complete 4-day recording.
- Optimize the random sampling algorithm for online learning and real-time data transmission.
- Discuss the effect of lossy data reconstruction on seismic data processing, e.g., cross-correlation.

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. *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). https://doi.org/10.1029/2021JC018375*