technische universität
dortmund

Wavelet-Based Clustering of DAS Data

Samuel Gende

September 25, 2024

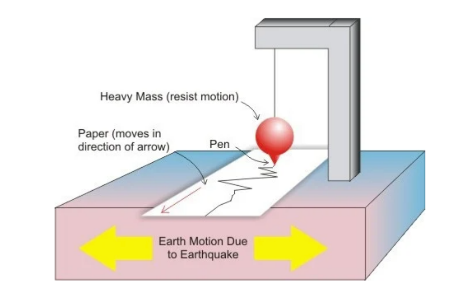
TU Dortmund University - Department of Computer Science

Overview

- Seismic Monitoring & DAS
- Wavelet-Based Clustering Technique & Changes
- Experimental Results
- Conclusion

Seismic Monitoring

- Seismic waves occur from sources such as Earthquakes, explosions, glaciers, ocean waves, traffic
- Continuously record ground motion with seismometers and geophones

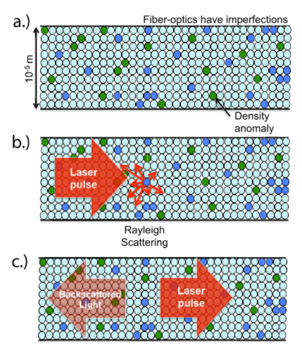


[1]

[1] AZoSensors. (2022, November 7). Application of seismometers in the measurement of earthquakes. `https://www.azosensors.com/article.aspx?ArticleID=7`

Distributed Acoustic Sensing (DAS)

- Uses fiber optic cables and laser light to measure strain caused by ground motion
- Cable is laid underground in a conduit and we measure ground motion at certain intervals
- Intervals at which the strain is measured are known as channels
- Interrogator sends laser light pulses then uses the backscattered light to calculate the strain



[2]

[2] Lindsey, N. J., Rademacher, H., & Ajo-Franklin, J. B. (2020). On the Broadband Instrument Response of Fiber-Optic DAS Arrays. Journal of Geophysical Research Solid Earth, 125(2). https://doi.org/10.1029/2019jb018145

Distributed Acoustic Sensing (DAS): Measuring Strain

- We Calculate the strain at a channel x_j by measuring the phase shift of backscattered light
- λ: Wavelength of the light used in the fiber-optic system.
- n: Refractive index of the fiber.
- L_G: Gauge length, the spatial resolution of the DAS system.
- C: Calibration constant for the DAS system.
- $\Delta\Phi$: Measured phase shift of the backscattered light.

$$e_{xx}(t, x_j) = \frac{\lambda}{4\pi n L_G C}\Delta\Phi. \tag{1}$$

[3]

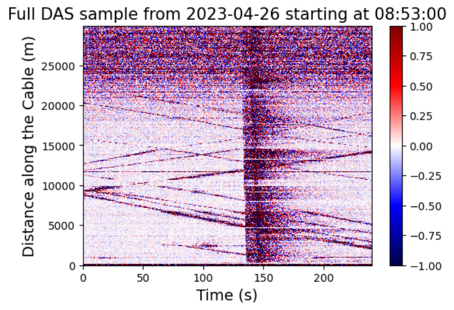
[3] Lindsey, N. J., Rademacher, H., & Ajo-Franklin, J. B. (2020). On the Broadband Instrument Response of Fiber-Optic DAS Arrays. Journal of Geophysical Research Solid Earth, 125(2). https://doi.org/10.1029/2019jb018145

Our DAS Array

- Located in Haast, South Island New Zealand
- Array Crosses the Alpine Fault pictured in red

4

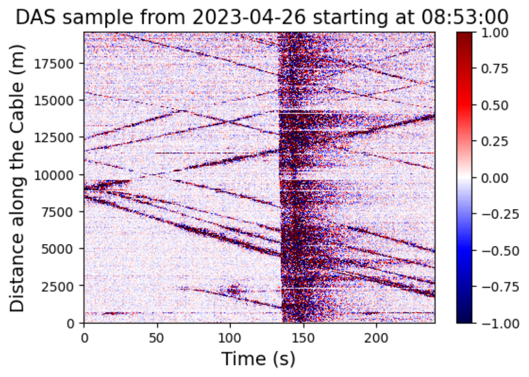
- 30km Array with 4m spacing
- Array has 7488 channels we use only 2449 with 8m channel spacing



[4] Konietzny, S., Lai, V. H., Miller, M. S., Townend, J., & Harmeling, S. (2024). Unsupervised Coherent Noise Removal from Seismological Distributed Acoustic Sensing Data. Authorea (Authorea). https://doi.org/10.22541/essoar.172253109.94701695/v1

Problems with DAS

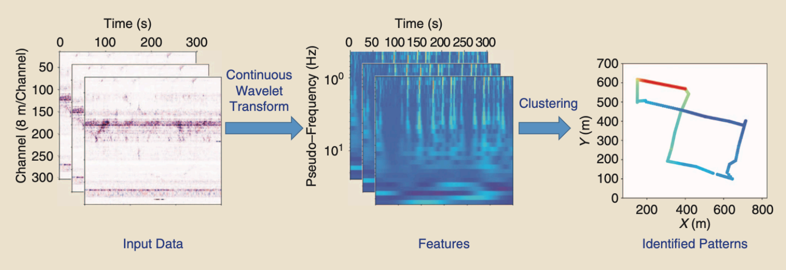
- Lower signal-to-noise ratio than seismographs and geophones
- Much more data demands automated processing
- **coherent noise from vehicles**



Solution: Wavelet-Based Clustering

- **Step 1:** Transform data into Features by applying a Continuous Wavelet Transform (CWT) to the DAS data.
- **Step 2:** Use k-means clustering to identify clusters associated with car noise.
- **Step 3:** Mute the identified car noise clusters.

- Technique is described in "A Seismic Shift in Scalable Acquisition Demands New Processing: Fiber-Optic Seismic Signal Retrieval in Urban Areas with Unsupervised Learning for Coherent Noise Removal"



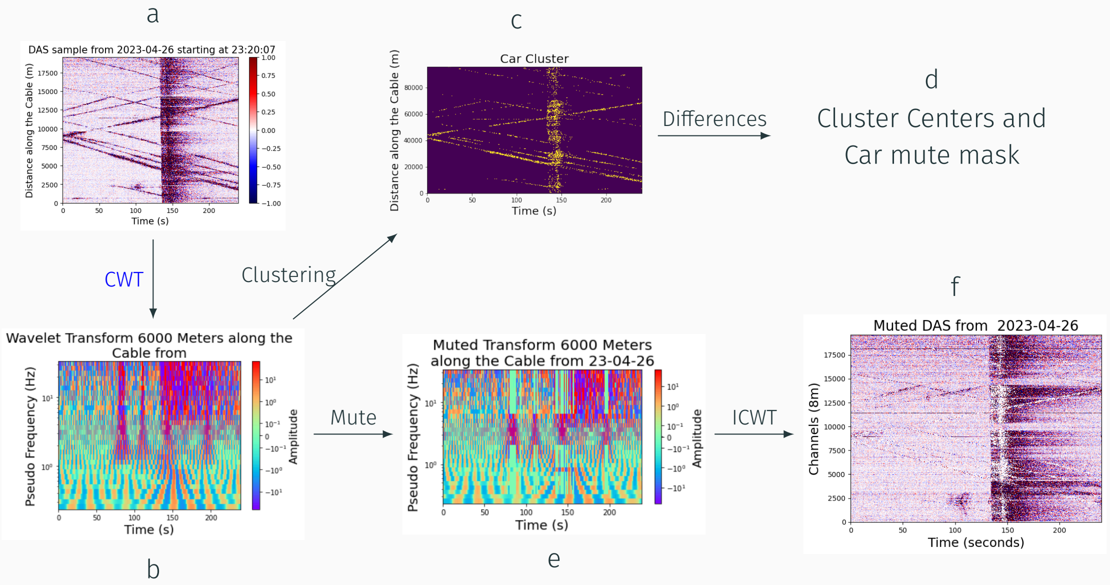
[5]

A Seismic Shift in Scalable Acquisition Demands New Processing: Fiber-Optic Seismic Signal Retrieval in Urban Areas with Unsupervised Learning for Coherent Noise Removal. (2018, March 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/8310692

Goals of the Thesis

- Re implement Martin et al.'s method
- Apply Wavelet-based clustering to Haast data
- Show that earthquake signals are preserved when muting car noise
- Test how to improve Martin et al.'s method

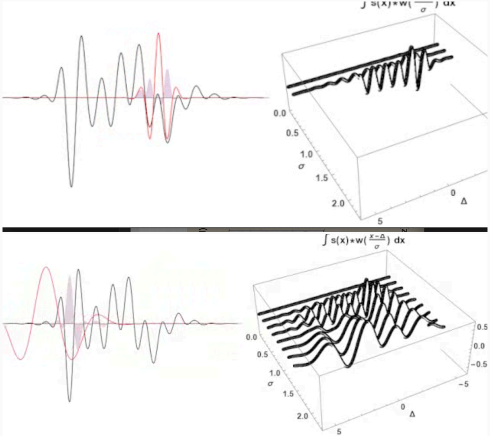
Our Wavelet-Based Clustering



Continuous Wavelet Transform

- Convolute Wavelet with Signal
- Evaluate CWT with different s and τ to scale and shift the wavelet
- Scaling the wavelet gives different time-frequency resolution
- Shifting the wavelet gives temporal information

$$CWT(\tau, s) = \frac{1}{\sqrt{|s|}} \int f(t)\psi * (\frac{t - \tau}{s})$$



[6]

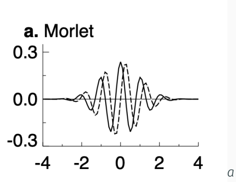
---

[6] Wikipedia contributors. (2024, July 19). Continuous wavelet transform. Wikipedia.

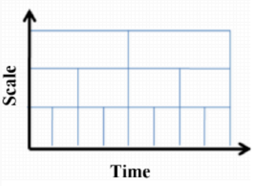
Our Wavelet Transform

- Discrete approximation of the CWT
- Use the complex Morlet wavelet results in complex scalogram
- Fourier Transform the signal and wavelet to speed up computation
- Inverse of CWT

$$x_t = C \sum_{j=0}^{J} \frac{R[CWT_t(s_j)]}{\sqrt{s_j}}$$



[a]



[b]

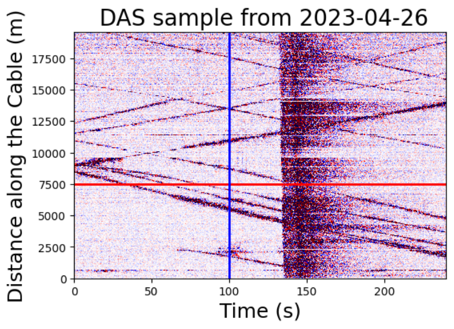
---

[a] Christopher Torrence and Gilbert P. Compo. **"A Practical Guide to Wavelet Analysis".** In: *Bulletin of the American Meteorological Society* 79.1 (1998), pp. 61–78. DOI: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/bams/79/1/1520-0477_1998_079_0061_apgtwa_2_0_co_2.xml.

[b] Machorro2023.

Applying CWT to DAS

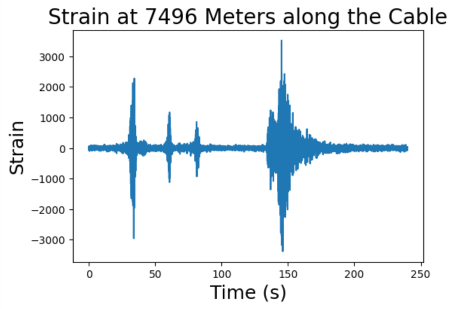
- Perform a CWT at every channel across time (Time CWT red)
- Perform a CWT at every time sample across the channels (Space CWT blue)

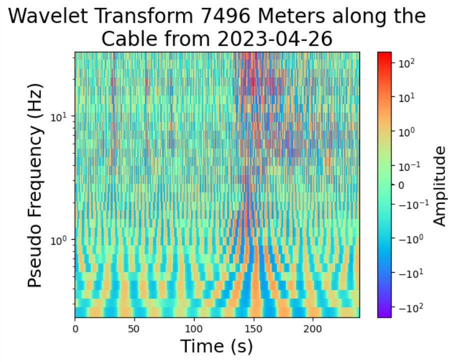


Applying CWT to DAS

- Original Signal at channel 937 has an earthquake event around 150 seconds
- Other peaks before 100 seconds due to cars

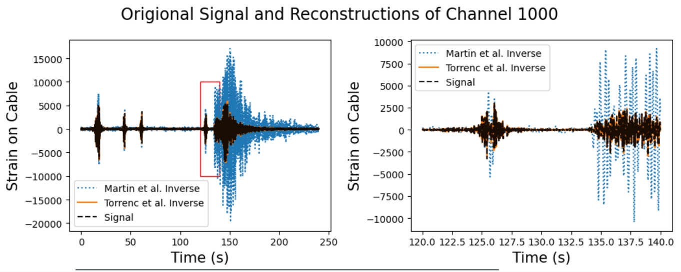
- We CWT the signal with 30 scales from 0.2 to 24 times mother wavelength
- Earthquake Signal and car signal still visible within CWT

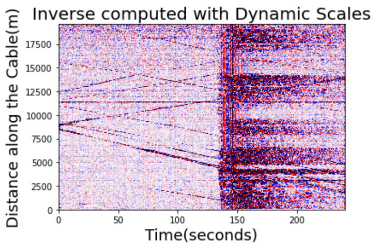




Change: Better scales

- Attempt to use scales from Torrence et al.[7] to get perform a more accurate reconstruction.
- Newer scales had better accuracy for one channel but caused problems for entire samples



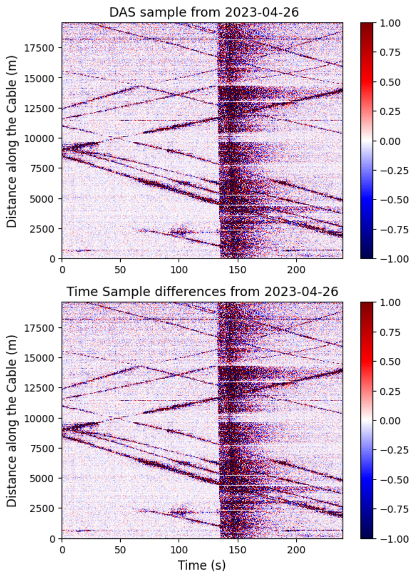


[7] Torrence, C., Compo, G. P., & Program in Atmospheric and Oceanic Sciences, University of Colorado, Boulder, Colorado. (1998). A practical Guide to Wavelet analysis. In Bulletin of the American Meteorological Society (Vol. 79, Issue 1, pp. 61–62).

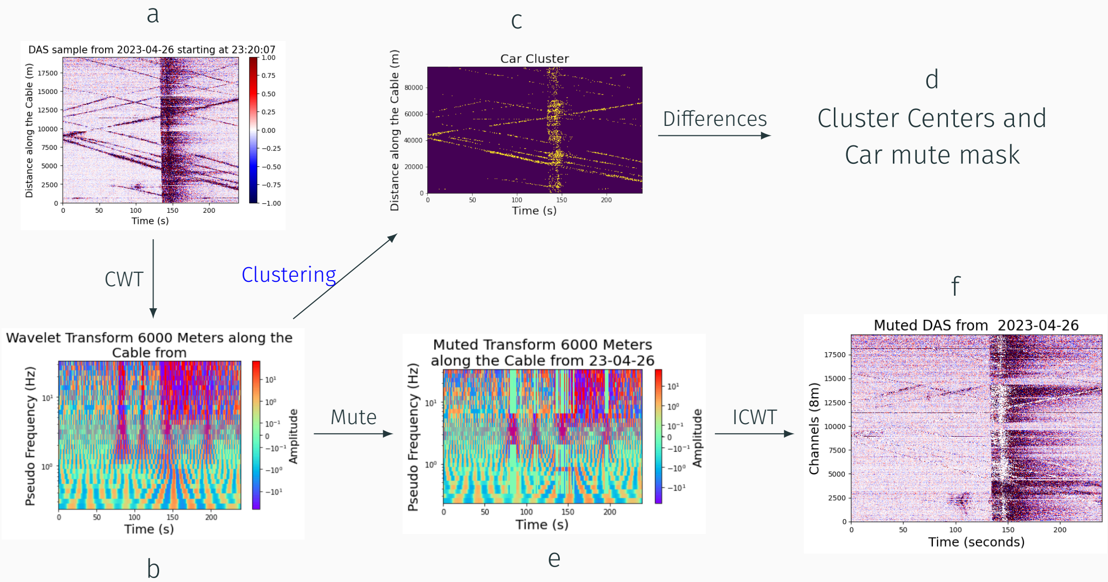
Change: Time Difference

- Martin et al. take difference between neighboring time samples before CWT
- This means inverse CWT is no the original signal, but the differences
- Time differences resulted in better detection of cars
- Earthquake signals and cars are still visible in time-difference plots

$$x_{j,t} = x_{j,t} - x_{j,t+1} \tag{2}$$

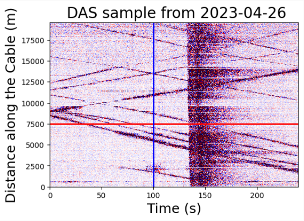


Our Wavelet-Based Clustering

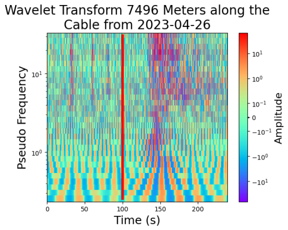


Features for Clustering

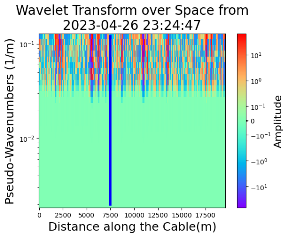
a



b



c

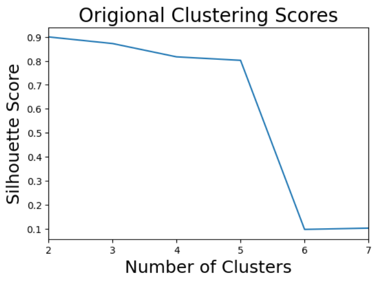


$$\text{Features} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$n = (n \text{ time scales} + n \text{ space}$

Clustering with k-means

- Transform 20 DAS samples into features
- Cluster these features with k-means
- use the silhouette score to settle on a good k



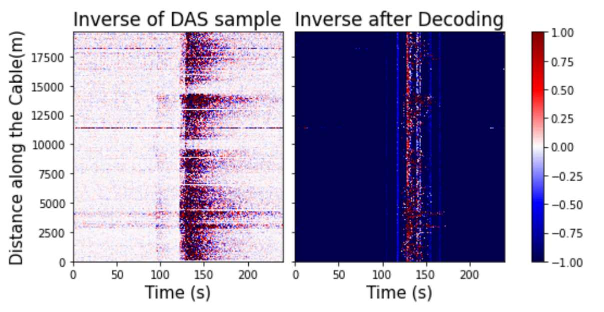
Change: Patched features

- Car and Earthquake signal differ in size of waves
- Idea is to use patches of DAS sample which should give more spatial information to clustering
- Group features into 2×2 patches of features then cluster patches
- Increases feature size by a factor of 4 which could be problamatic

Change: Dimensionality reduction with Autoencoder

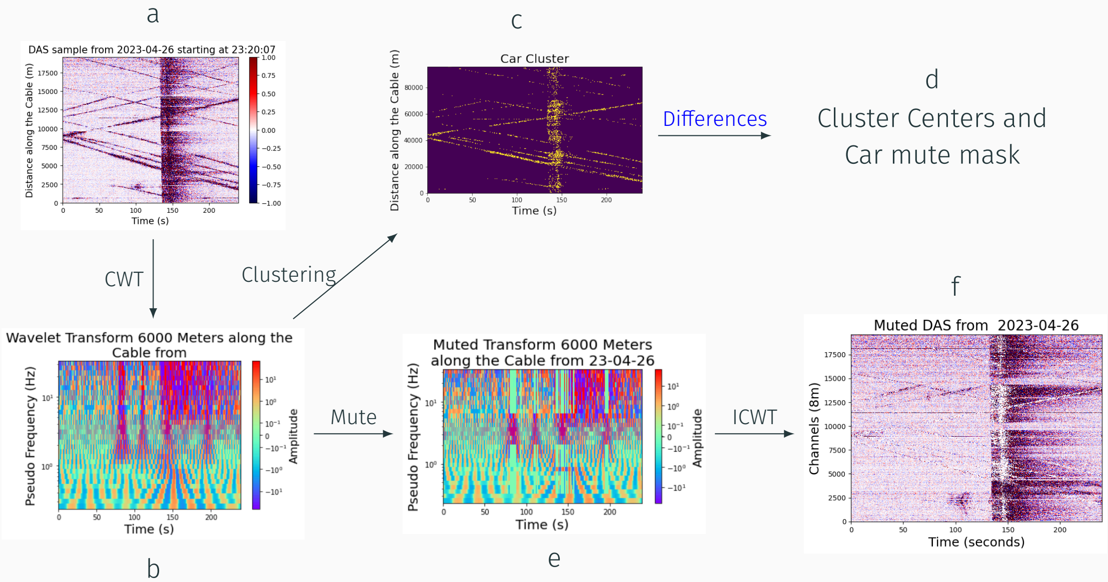
- k-means struggles with high-dimension clustering
- use an autoencoder to learn a lower dimension latent space
- cluster in low dimension latent space
- tested latent space clustering with MNIST [8]

Method	Dimensions	Accuracy
Flatten	(784,)	46.71%
PCA	(10,)	48%
DAC	(10,)	58%



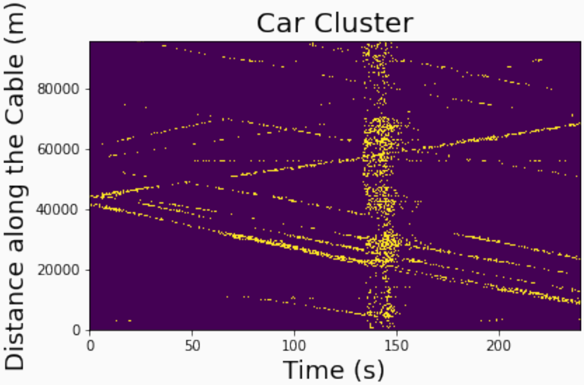
[8] Lu, S., & Li, R. (2021, February 15). DAC: Deep autoencoder-based Clustering, a general deep learning framework of representation learning. arXiv.org. https://arxiv.org/abs/2102.07472

Our Wavelet-Based Clustering



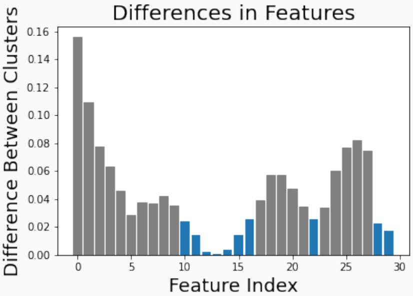
Identifying Car Clusters

- Decide which cluster or clusters correspond to car noise
- smaller clusters generally where more noisy
- Plotting the clusters on the DAS for final decision of car clusters
- k=5 with 3 car clusters worked best

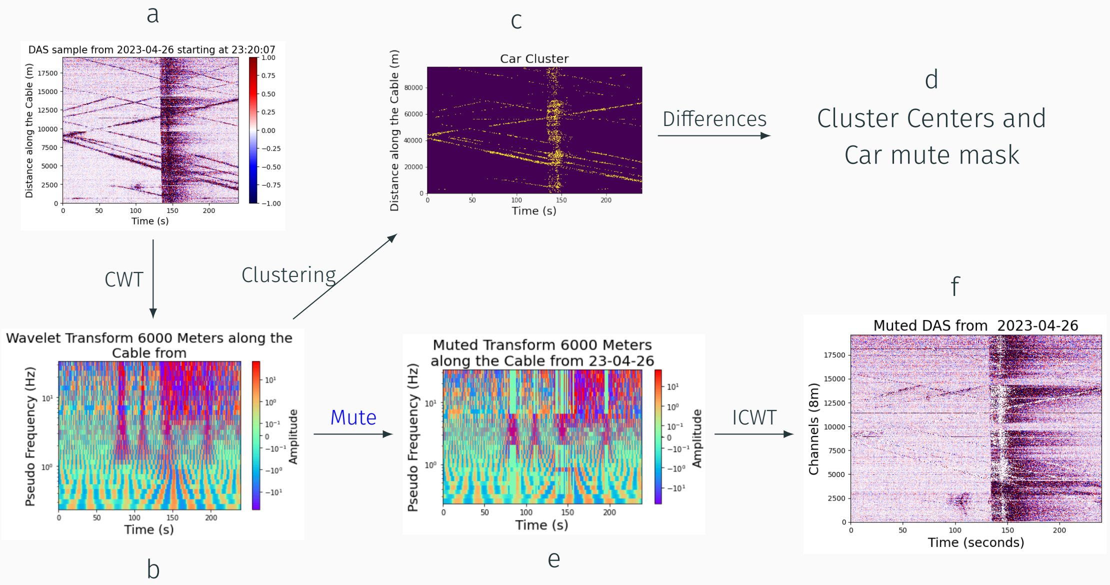


Car Mute Mask

- Once car noise can be identified we need to remove noise
- We do this with mute mask which is a binary mask where some scales are multiplied with zero
- Preserve the earthquake signal by only muting 20 scales that differentiate car noise from other signals
- We only mute time scales since they are used in inverse CWT

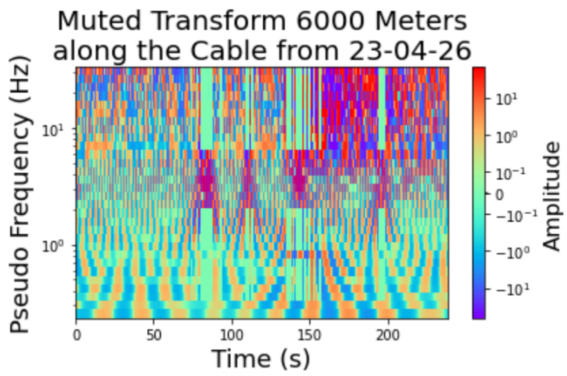


Our Wavelet-Based Clustering

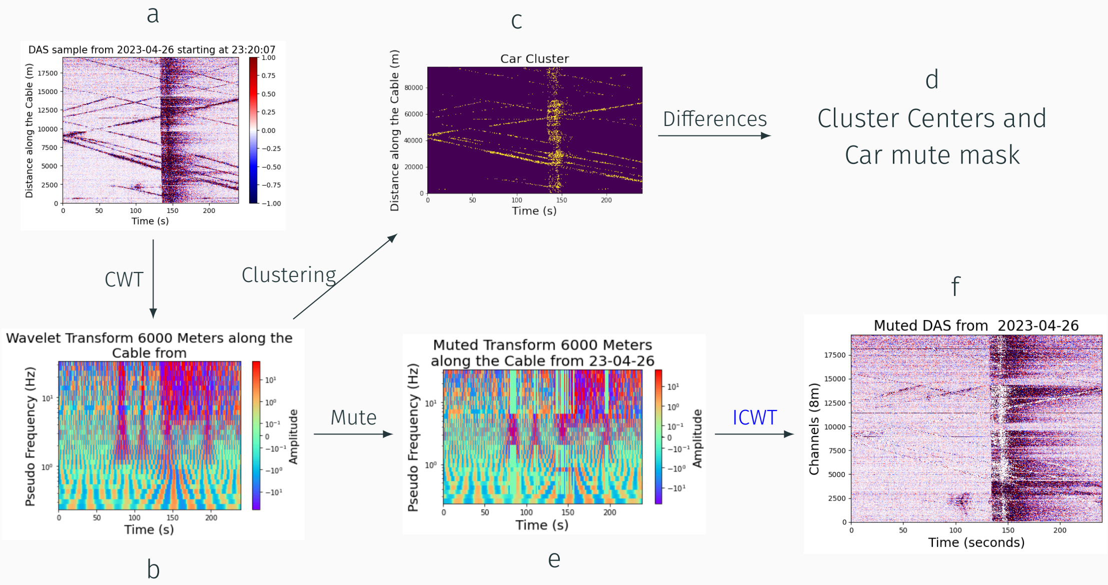


Applying Car Mute to new Samples

- **Step 1:** Transform new sample into features
- **Step 2:** Classify features with the distance to cluster centers
- **Step 3:** Features classified as car noise are multiplied with car mute mask



Our Wavelet-Based Clustering



Inverse CWT of DAS Samples

- Perform inverse CWT over time at every channel
- Inverse of the muted samples means we won't reconstruct the car noise
- Reconstructed samples should have less car noise and increase local SNR around earthquake signal

Results

Estimating the SNR with Semblance

- We can Estimate local SNR with just the noisy DAS sample
- Uses the correlation of channels to estimate local SNR with semblance [9]
- Zero lag cross correlation of two channels d_k, d_l :

$$R_{d_k d_l} = \sum_{i=1}^{N} d_{ik} d_{il}$$

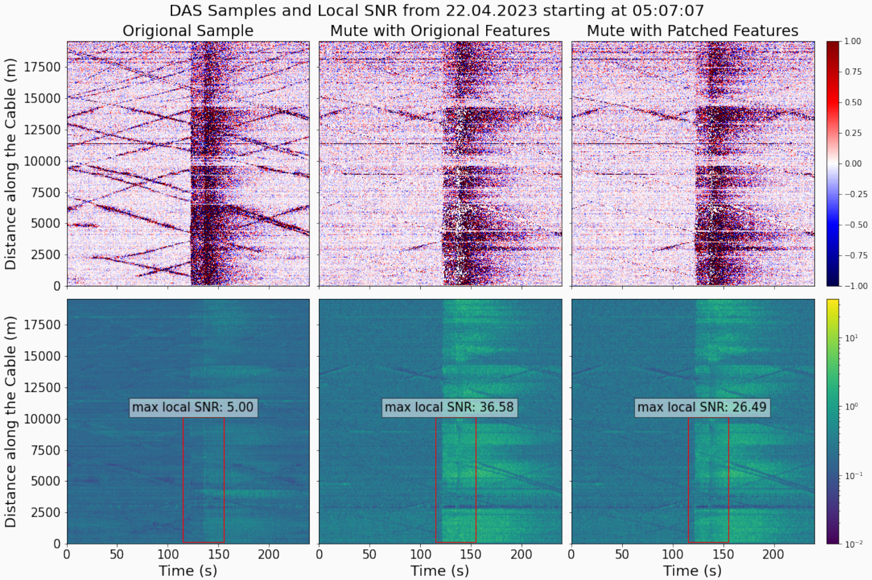
- Semblance:

$$S = \frac{\sum_{k=1}^{M} \sum_{l=1}^{M} R_{d_k d_l}}{M \sum_{l=1}^{M} R_{d_l d_l}}$$

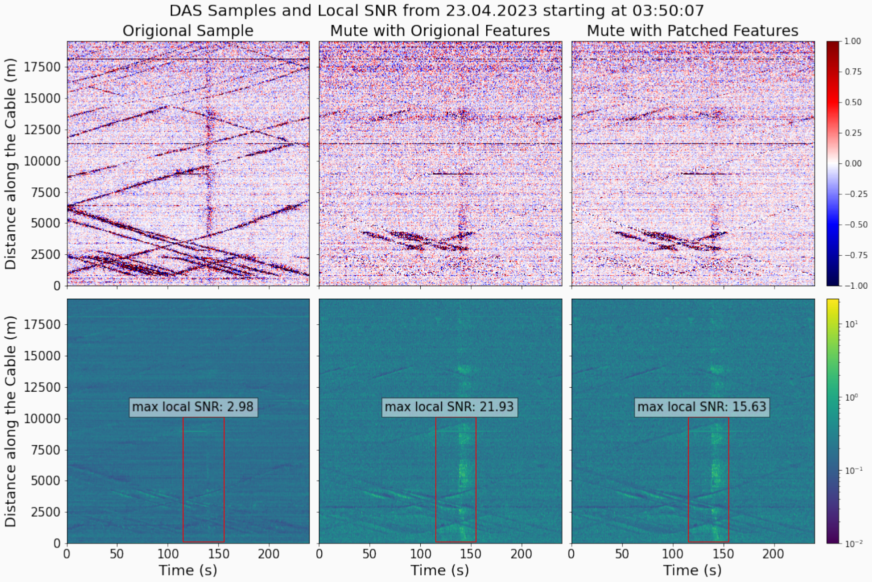
$$SNR_{sem} = \frac{S}{1 - S}$$

[9] Lapins, S., Butcher, A., Kendall, J., Hudson, T. S., Stork, A. L., Werner, M. J., Gunning, J., & Brisbourne, A. M. (2023). DAS-N2N: machine learning distributed acoustic sensing (DAS) signal denoising without clean data. Geophysical Journal International, 236(2), 1026–1041. https://doi.org/10.1093/gji/ggad460

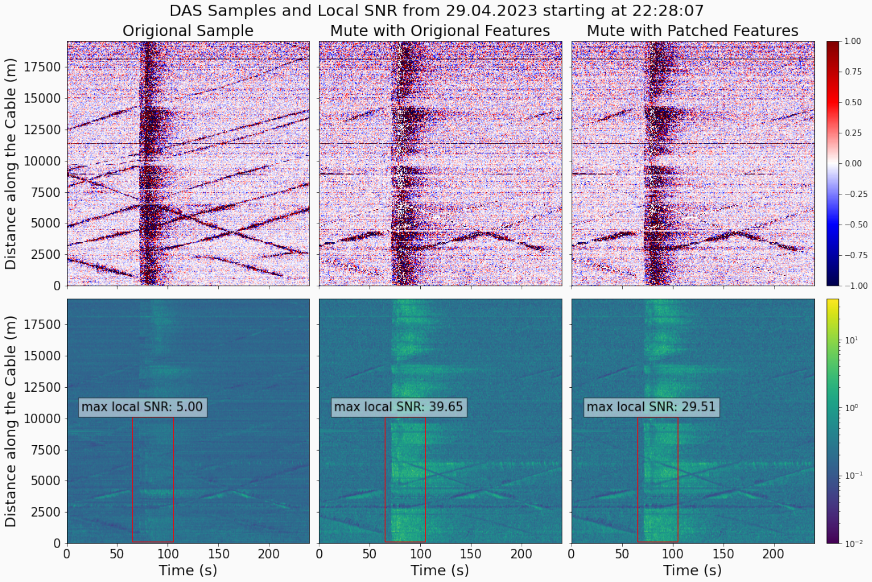
Results of DAS sample from 22.04.2023



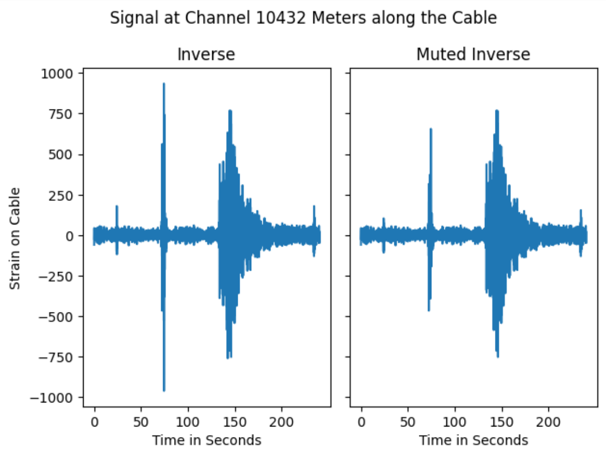
Results of DAS sample from 23.04.2023



Results of DAS sample from 29.04.2023



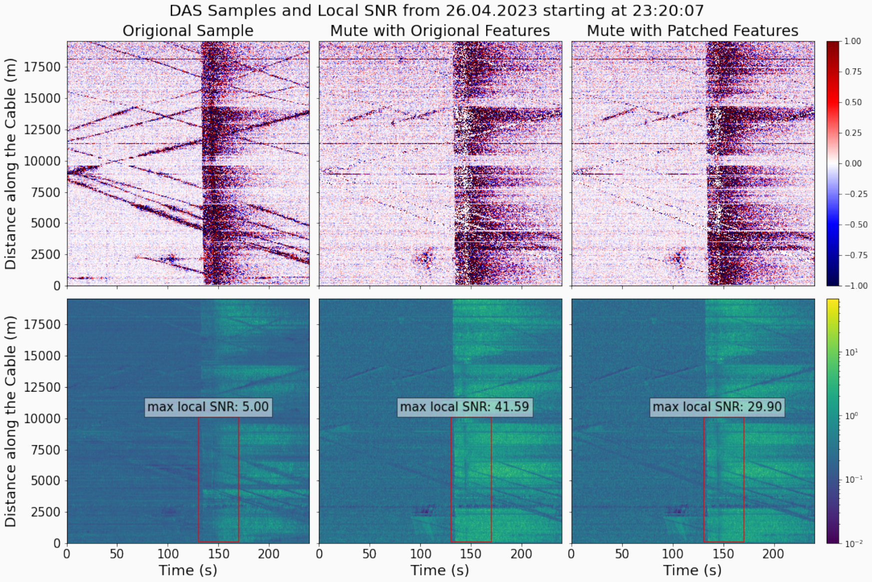
Results of a Single Channel



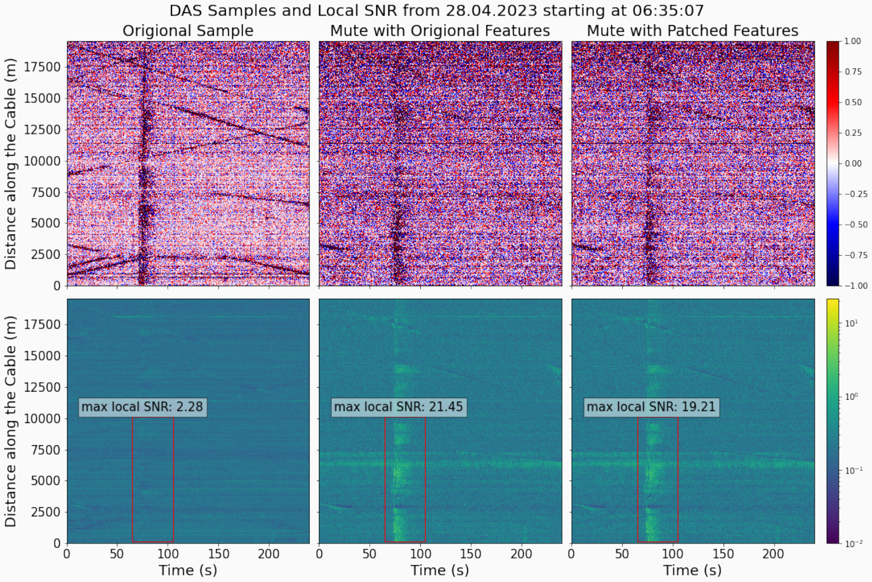
Conclusion

- Reimplemented Martin et al.s Wavelet-Based Clustering
- Tested Wavelet-based clustering on Haast data
- Found that Wavelet-based clustering increased local SNR
- Tested larger patched features

Results of DAS sample from 23.04.2023



Results of DAS sample from 28.04.2023



Results of DAS sample from 21.04.2023