Summarizing Low-dimensional Patterns in Long-term Echosounder Time Series from the U.S. Ocean Observatories Initiative Network

Wu-Jung Lee, Applied Physics Laboratory

Valentina Staneva, eScience Institute

University of Washington, Seattle, USA

April 25, 2022

Wu-Jung Lee
Acoustical Oceanographer

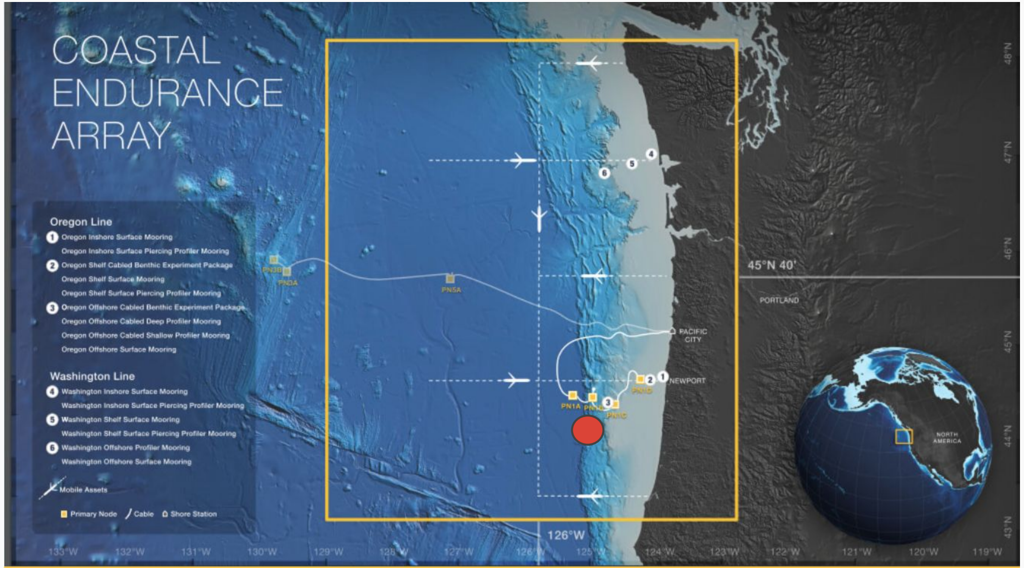
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Applied Physics Laboratory
UNIVERSITY of WASHINGTON

Motivation

OCEAN
OBSERVATORIES
INITIATIVE

- ➢ Continuous (24/7) moored data collection
- ➢ Commissioned since 2015
- ➢ Committed to 25+ years of operation
- ➢ Multiple instruments simultaneously sensing the environment
- ➢ Upward looking echosounders (200m)



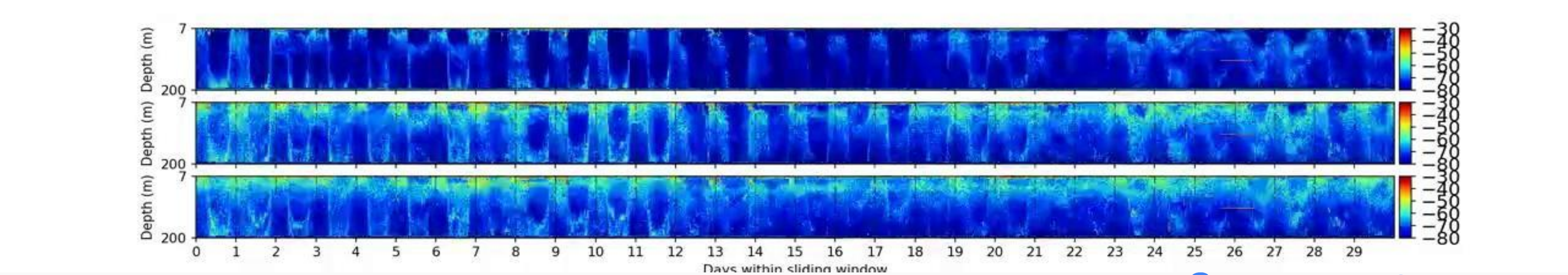
Credit: Center for Environmental Visualization, University of Washington

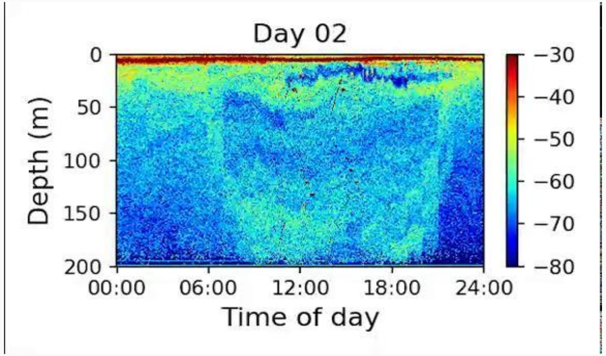
How can we extract information from these long-term time series?

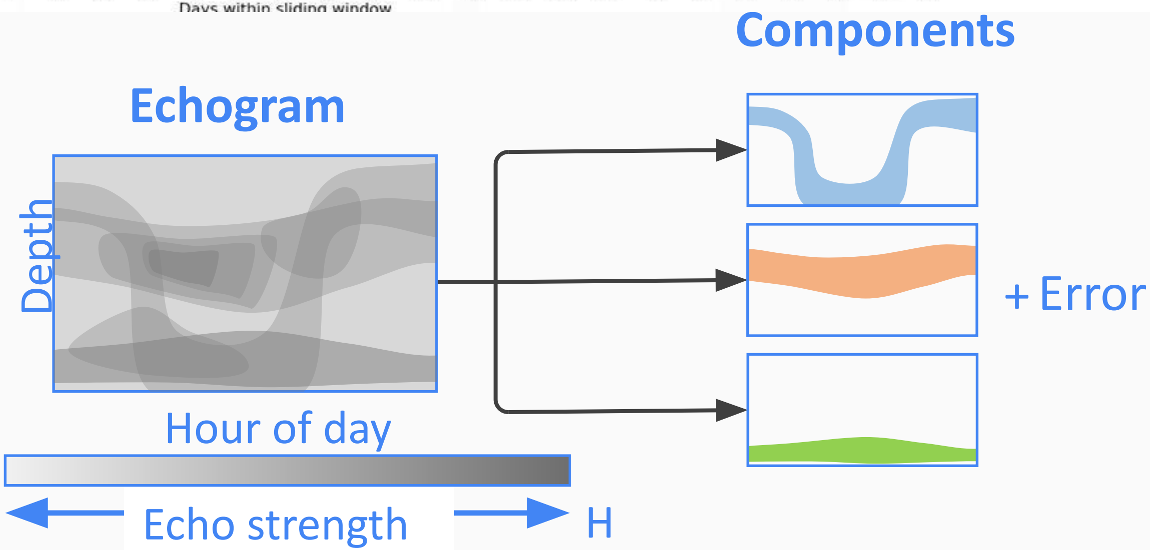
Low-dimensional Representation of Echograms

Echogram: 12 months of data

Depth (m)







Components

Echogram

Depth

Hour of day

+ Error

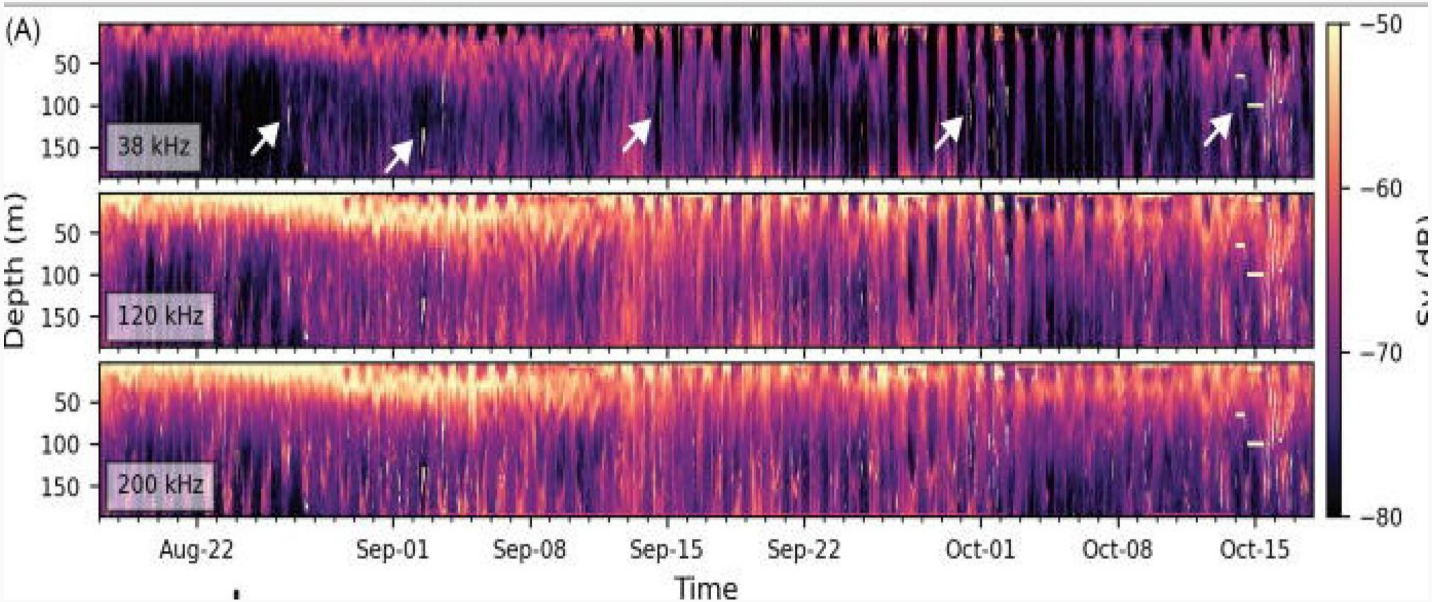
L

Echo strength

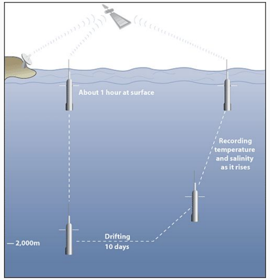
H

Data Outliers

2 months of data: Aug 17- Oct 17, 2015



Ocean profiler



Decomposition analysis (such as Principal Component Analysis) on data with outliers yields a corrupted result, as the magnitude of the outliers can dominate the cost.

Robust Principal Component Pursuit for Outlier Removal

$$\min_{L,S} \|L\|_* + \gamma\|S\|_1,$$
$$\text{subject to } L + S = X$$

Low rank
(Sonar Patterns)

Sparse
(Profiler Artifacts)

Rank: $rank(L) = \#\,\text{nonzero singular values } (\sigma_i)$

Nuclear Norm: $\|L\|_* = \sum_i \sigma_i(L)$

Entrywise L1-norm: $\|S\|_1 = \sum_{ij} |s_{ij}|$

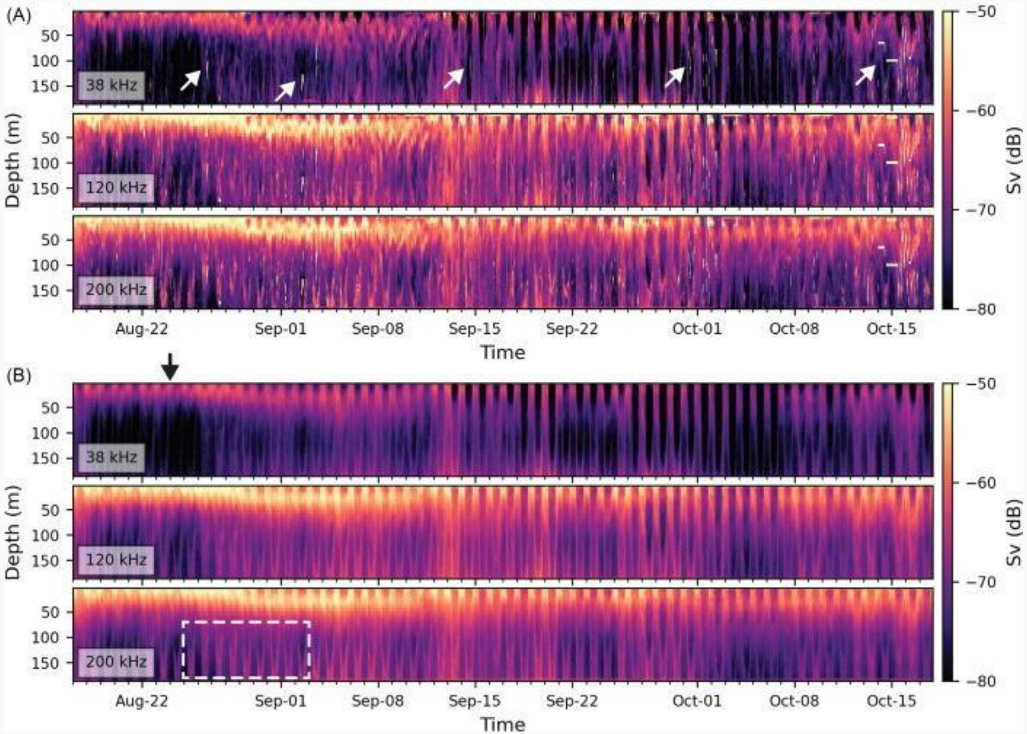
If such a decomposition exists the solution can be found ***exactly,*** *no tuning of* γ *needed!*

- Candes'09 Robust Principal Component Analysis?

Robust Principal Component Pursuit for Outlier Removal

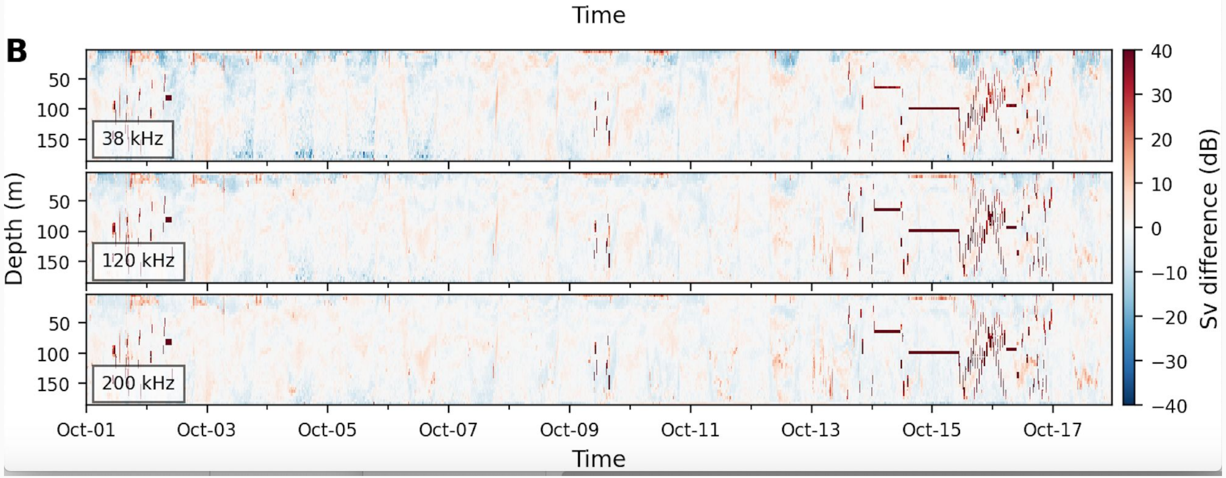
Raw Data *(X)*

Low Rank Component *(L)*
Denoised Data



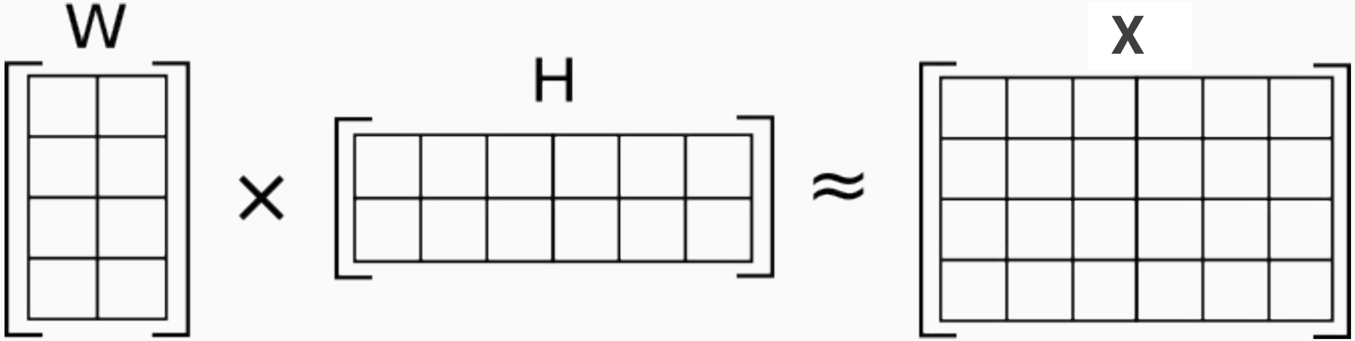
Robust Principal Component Pursuit for Outlier Removal

Sparse Component (*S*):



- ➢ Some signal is in the sparse component.
- ➢ There is intrinsic variation in the low rank patterns over time.
- ➢ Not crucial for extracting dominant patterns.

Nonnegative Matrix Factorization (NMF) for Echogram Pattern Discovery



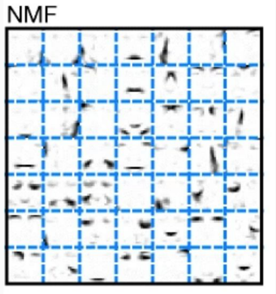
Latent Variables/ Components

Weights/Activations

Data

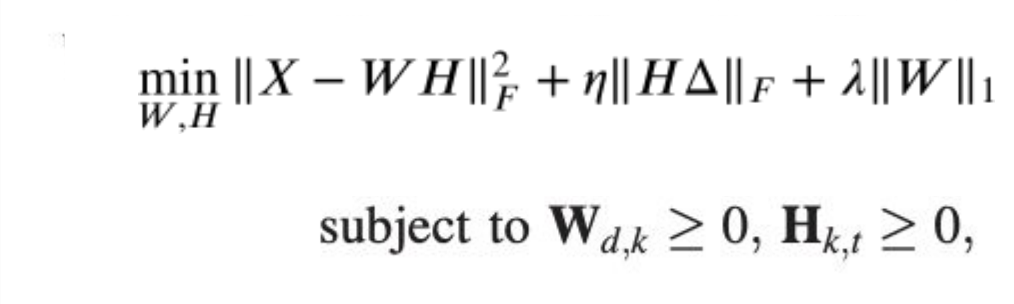
$$\min_{W,H} \|X - WH\|_F^2,$$
$$\text{subject to } W \geq 0,\ H \geq 0$$

- Total backscatter is built of backscatter of individual components
- Lee et.al., Learning the parts of objects by nonnegative matrix factorization



Temporally Smooth NMF

Smooth Activations Sparseness

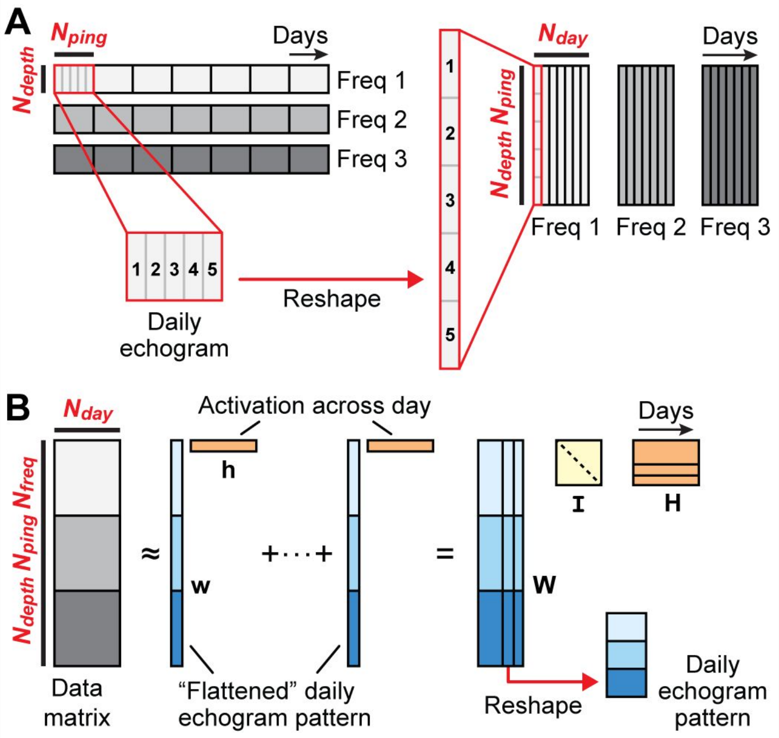


$$\min_{W,H} \|X - WH\|_F^2 + \eta\|H\Delta\|_F + \lambda\|W\|_1$$

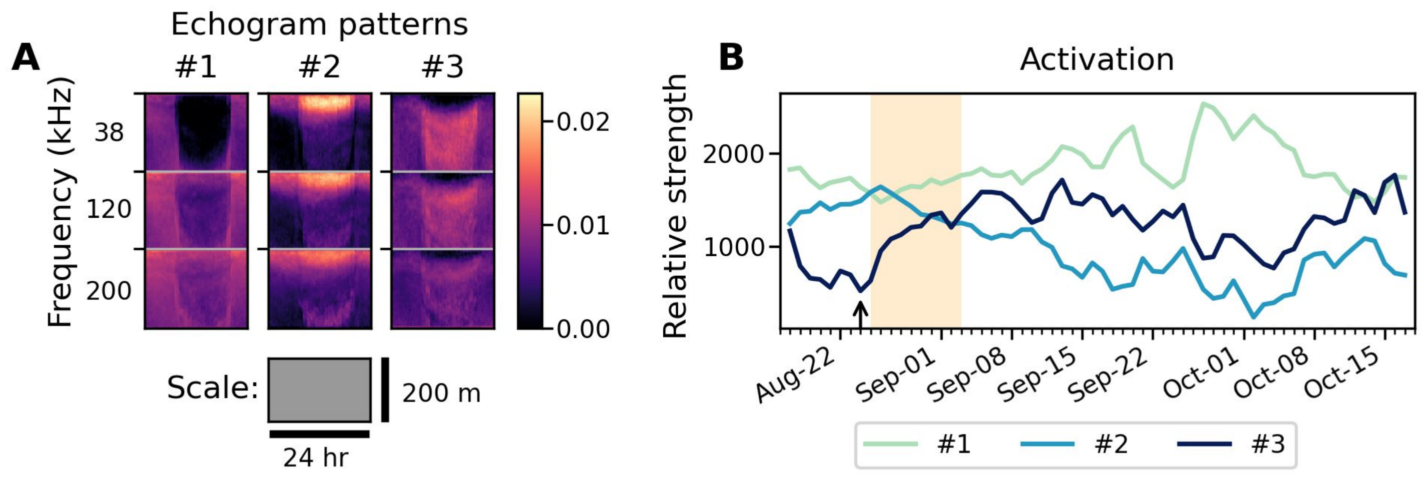
$$\text{subject to } \mathbf{W}_{d,k} \geq 0,\ \mathbf{H}_{k,t} \geq 0,$$

- ❏ [Fabregat et. al., solving NMF with smoothness and sparsity constraints]
- ❏ Python Package: https://pypi.org/project/time-series-nmf/

Sonar Data Reorganization for Matrix Decomposition



NMF Results



#1: DVM (zooplankton-like)
#2: Subsurface Layer (zooplankton-like)
#3: fish-like

In a Nutshell

- ➢ **Robust PCA** is powerful for automatically removing outliers.
- ➢ **Nonnegative Matrix Factorization** discovers biologically meaningful temporal patterns.

More details:

- ■ Lee W.-J., Staneva V., Compact representation of temporal processes in echosounder time series via matrix decomposition

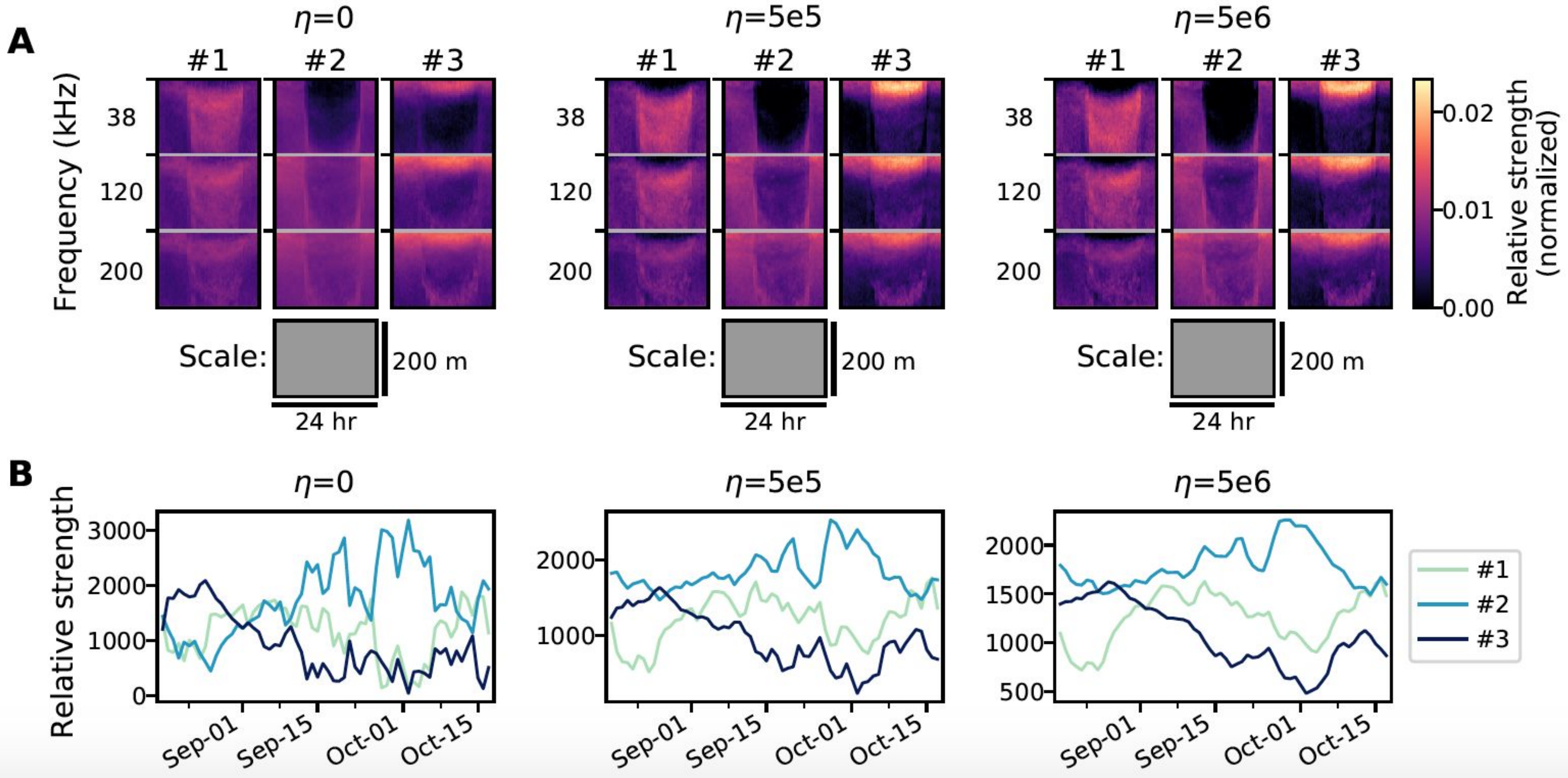
Ongoing and Future Work:

- ➢ Expand to years of Ocean Observatories Initiative data
- ➢ Analyze with conjunction with other environmental variables

NSF
GORDON AND BETTY MOORE FOUNDATION
ALFRED P. SLOAN FOUNDATION

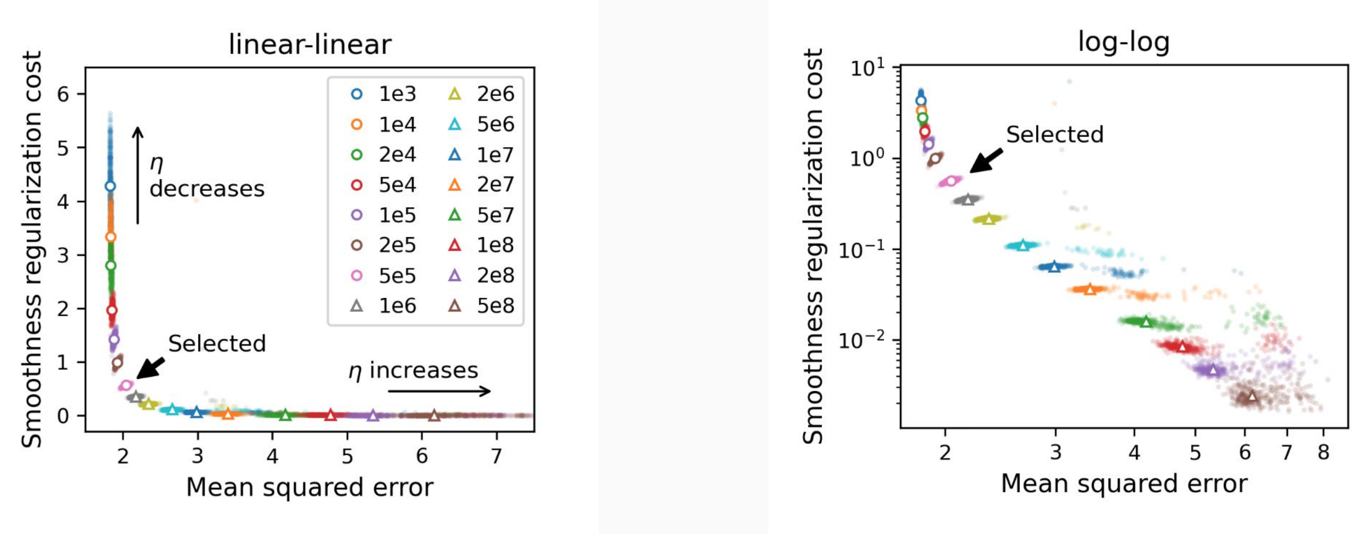
Extra

Effect of Smoothing Parameter



L-curve Method for Selecting Regularization Parameters

Trade-off between minimization cost and smoothness.



(point cloud correspond to multiple runs)

[Oraintara, S., et. al., A method for choosing the regularization parameter in generalized Tikhonov regularized linear inverse problems]

Robust Principal Component Pursuit for Outlier and **Noise** Removal

(Zhou et.al, Stable Principal Component Pursuit)

Instead Assume:

$$X = \underset{\substack{\nearrow \\ \text{Low rank}}}{L} + \underset{\substack{\uparrow \\ \text{Sparse}}}{S} + \underset{\substack{\nwarrow \\ \text{Noise}}}{N}$$

Small noise: $\|N\|_F < \delta$*we can solve the following problem*

$$\min_{L,S} \|L\|_* + \gamma\|S\|_1,$$
$$\text{subject to } \|X - L - S - N\|_F < \delta$$

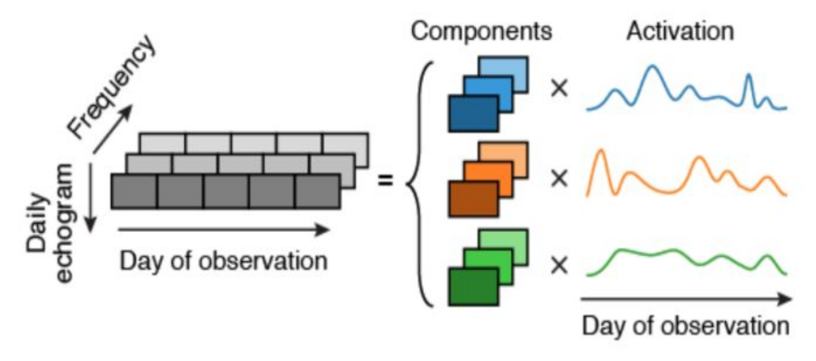
“Stable”: small noise => small reconstruction error

Extra parameter that needs to be tuned!

https://github.com/ShunChi100/RobustPCA/

Tensor Decomposition of Sonar Data

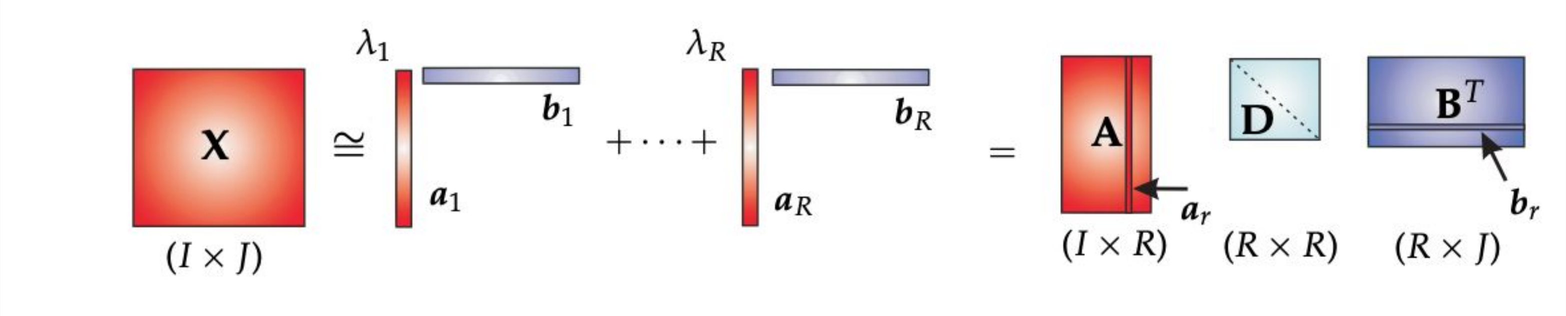
Treat each dimension separately?



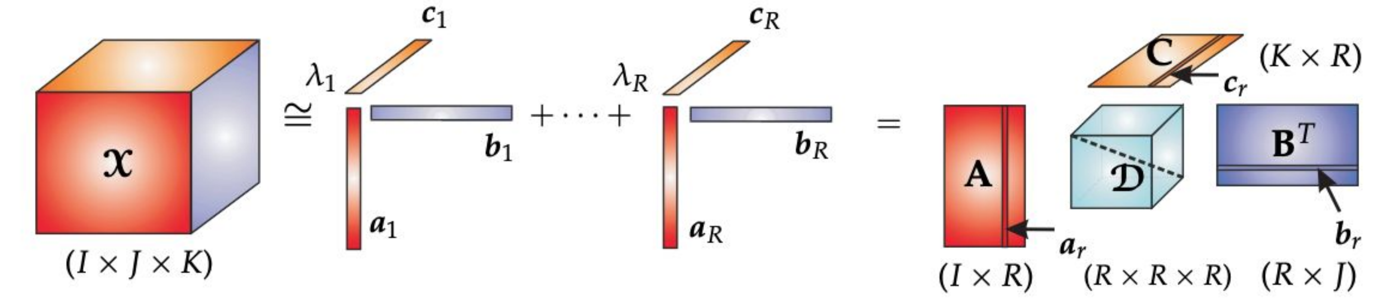
Matrix decomposition => Tensor Decomposition

Kruskal Tensor Decomposition

2D decomposition as a sum of outer products (SVD):



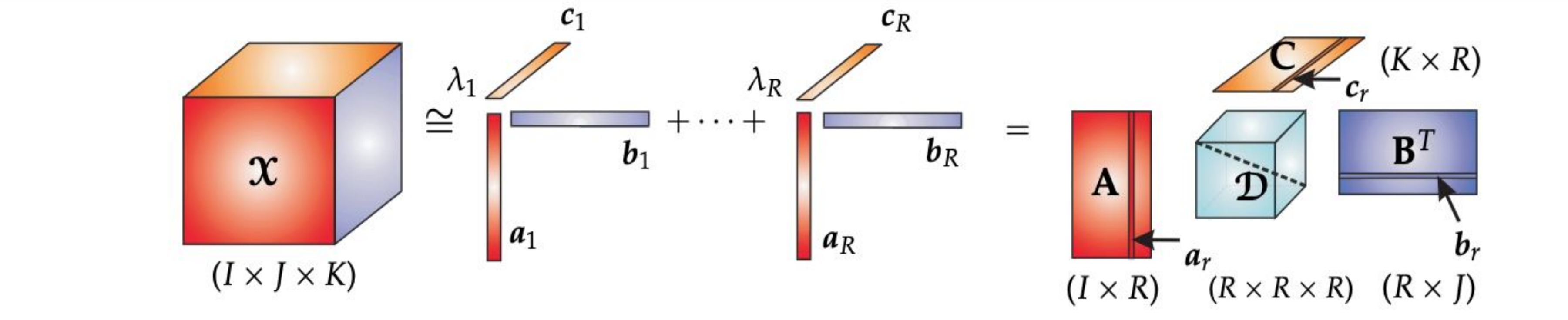
3D decomposition as a sum of outer products (higher order SVD)



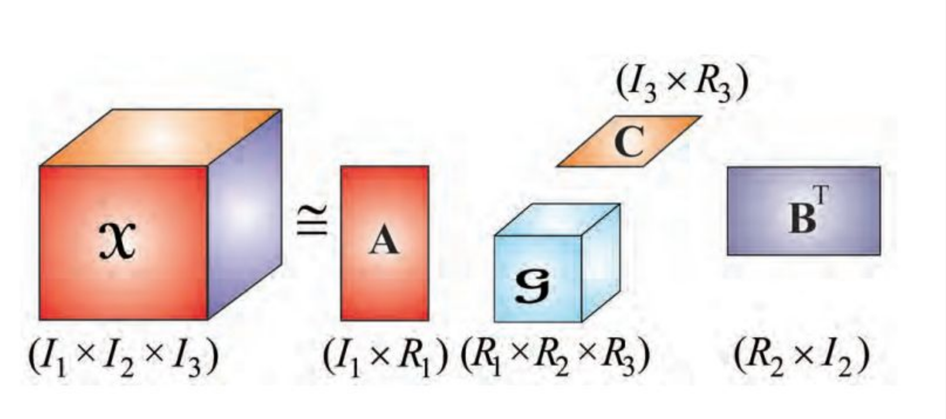
Cichocki, A. et. al., Tensor Decompositions for Signal Processing Applications: From Two-way to Multiway Component Analysis

Kruskal vs Tucker Tensor Decomposition

Kruskal Decomposition:



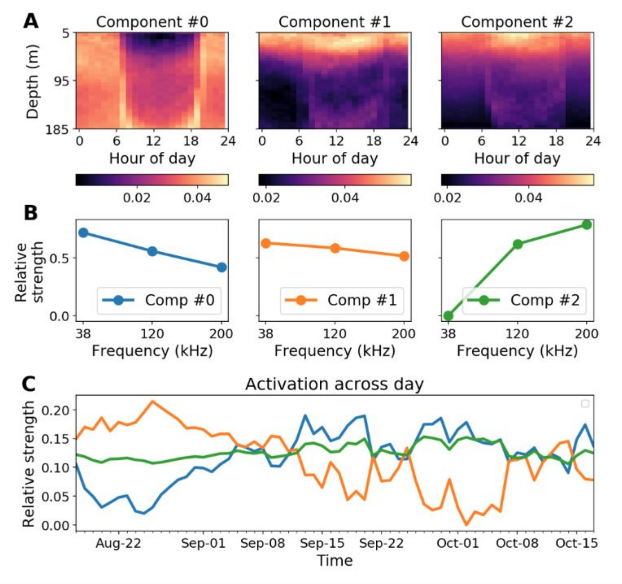
Tucker Decomposition:



Tensor Decomposition Results

Kruskal Nonnegative Tensor Decomposition

`tensorly` package



Automate heuristic methods based on thresholding rules about the frequency response which depend on correct calibration.

Tensor Decomposition Considerations

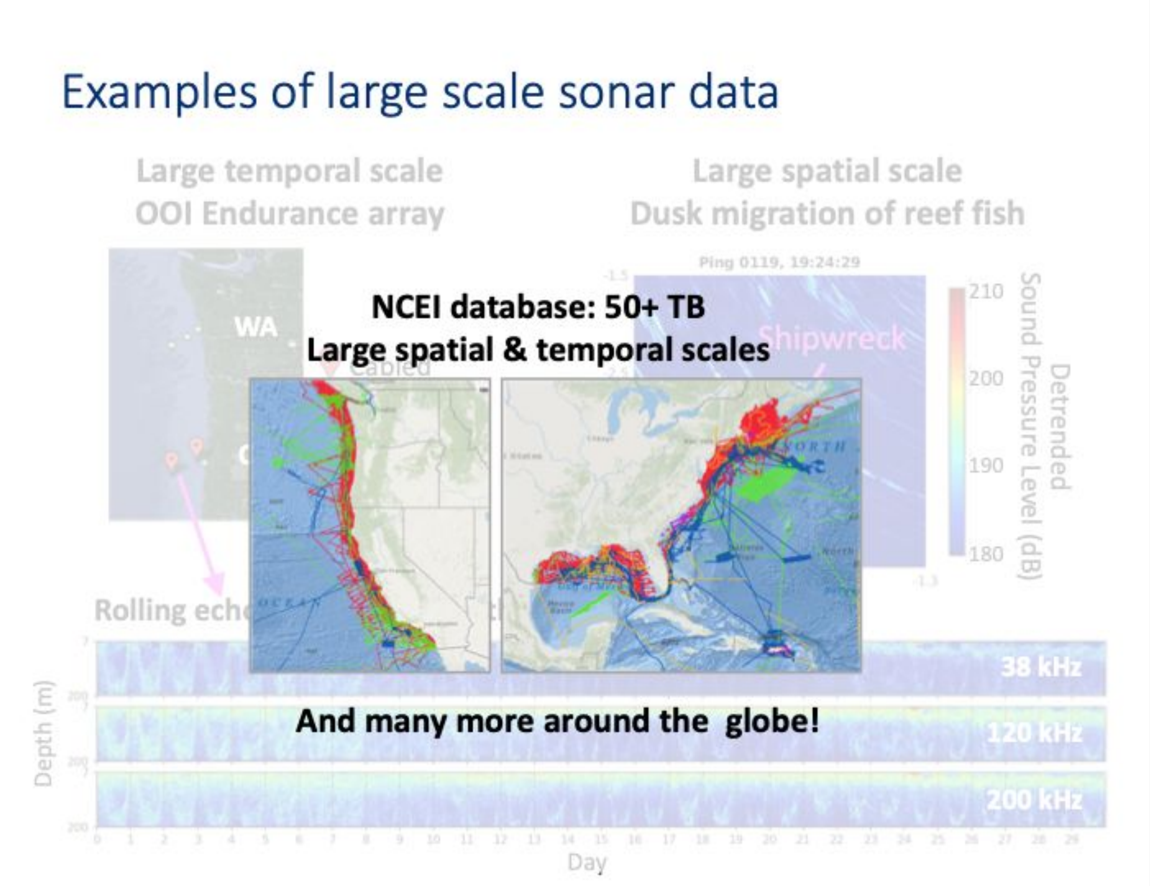
The rank-1 constraint of the Kruskal Decomposition is very limiting:

- ➢ Maximum 3 components (the lowest dimension)
- ➢ Components should have the same frequency response
- ➢ More noise in the reconstruction

The solution is almost certainly unique (as opposed to matrix decomposition)

Better suited for broadband data: many frequencies

Future: Many Large Datasets to Analyze



Examples of large scale sonar data

Large temporal scale
OOI Endurance array

Large spatial scale
Dusk migration of reef fish

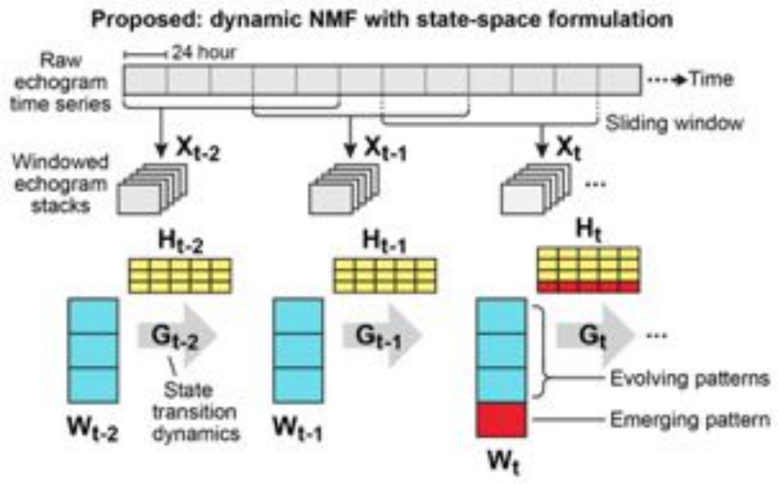
NCEI database: 50+ TB
Large spatial & temporal scales

And many more around the globe!

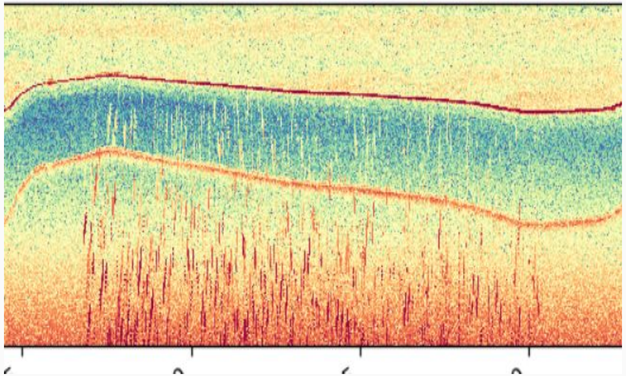
Future: Emerging, Evolving, and Fading Patterns

Within long time series the patterns are gradually changing over time:

➢ Add constraints on the patterns W (as opposed to the activations).



Future: Application to Ship Data

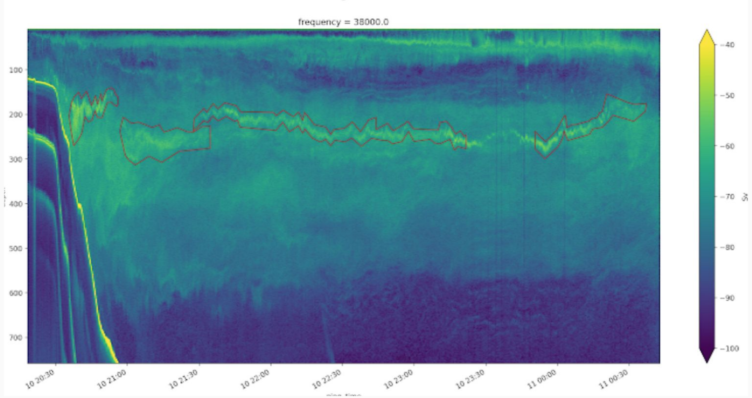


Ship is moving!

- ➢ Bottom variation
- ➢ Environment variation
- ➢ Ship speed variation

Future: Compare to Ground Truth

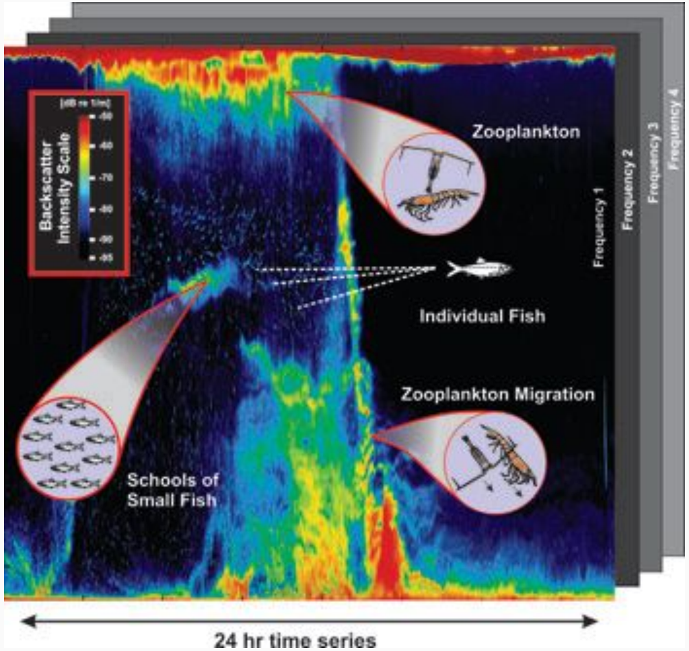
- Survey Cruise annotations
- Trawl data



Collaborating with NOAA Northwest Fisheries Science Center

Water Column Sonar Data

Data: depth, time, frequency



Acoustic Zooplankton Fish Profiler ™

Goals:

- ➢ Discover patterns of marine organism activities
    - ○ Distinguish between fish and zooplankton
    - ○ Estimate species abundance
    - ○ Detect migrational patterns
- ➢ Relate to physical processes and external phenomena