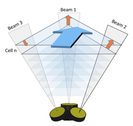
echopype: Toward Interoperable and Scalable Ocean Sonar Data Processing

Wu-Jung Lee[1], Kavin Nguyen[1], Valentina Staneva[2],
Applied Physics Laboratory[1], eScience Institute[2],
University of Washington

Motivation

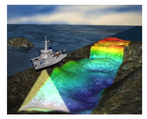
Sonar is widely used in ocean observation.



ADCP (Acoustic Doppler Current Profiler)

Sidescan sonar

Echosounder, fishfinder



Multibeam sonar mapping

Ocean observing systems are collecting terabytes of data.

Challenges:

Interoperability:

- Many manufacturers, many sonar models
- Proprietary software or open-source software written in proprietary languages (e.g., Matlab)

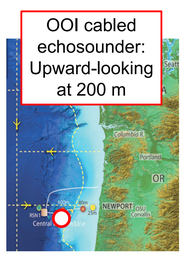
Scalability:

- Current analysis workflow is labor-intensive
- No support for parallel computation with random-access file formats

Reproducibility:

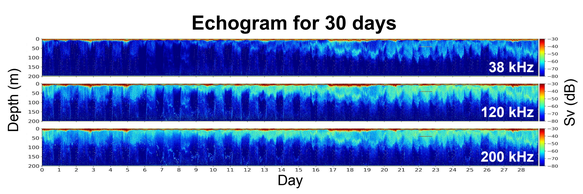
- Currently mostly GUI-based: good for exploration but hard to reproduce

Sonar Data Example



The cabled echosounder in the Ocean Observatories Initiative (OOI) Endurance Array off Oregon coast has been continuously streaming data almost real-time through the web since 2014. However, due to the proprietary format of the sonar data, they are currently not included in the web-portal and the OOI API, where data from most other instruments can be found.

Echogram for 30 days



The up- and down- daily patterns correspond to the Diel Vertical Migration (DVM), during which marine animals move toward the surface at night, and return to deeper during the day. There are also seasonal shifts in the composition layers.

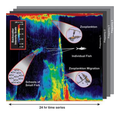


Image Source

Sonar Data Processing Workflows

SIMRAD, KONGSBERG, BioSonics, ASL Environmental Sciences, EdgeTech, NORTEK, ⋮

Data wrangling is exhausting and expensive.
We should be spending more time here!

Clean & combine data → echoview / MATLAB → Perform actual analysis → Fisheries & ecological interpretation

Current sonar processing workflows are labor intensive and researchers spend most of their time on data conversion between proprietary data formats and repetitive pre-processing steps. Further, the use of paid (and often GUI-based) packages hinders the scalability and reproducibility of the analyses, which prevents researchers from embarking on large scale ocean observational studies.

echopype

Simplify data wrangling, so that we can spend more time here!

echopype → Existing analyses (scikit-learn, TensorFlow, PYTORCH, ⋮) → Ecological & fisheries interpretation

xarray, DASK, netCDF

vs NumPy

- simple for Matlab users
- in-memory operations
- no labeled indexing or metadata support

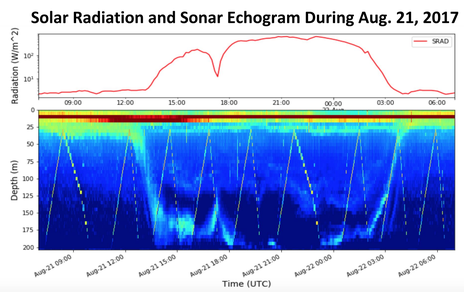
echopype is an open-source Python package which aims to address current challenges in sonar data processing by adopting a netCDF format and leveraging existing Python libraries for large scale data processing and visualization (such as *xarray* and *dask*).

- uses ICES SONAR-netCDF4 convention
- supports EK60, AZFP data format conversion
- allows random data access for easy time period and frequency selection
- preserves labels for scientific integration and interpretation
- performs calibration, denoising, mean volume backscattering strength calculation

Facilitating Scientific Data Integration

Watching a solar eclipse using an OOI sonar (with a few lines of code).

Access OOI data → Process using *echopype*
Access solar radiation data from National Data Buoy Center (NDBC) → Combine and visualize



We can directly pull data from different web repositories and align them in time. We see that there is an upward zooplankton movement corresponding to the dip in the solar radiation. The animals were fooled by the solar eclipse and started preparing for the night! (Binder Notebook)

Future Directions

Format Support

- Currently supports EK60, AZFP data conversion
- Expand to EK80, ADCP, and other sonar systems

Cloud Support

- Test scalability on a kubernetes cluster
- Deploy example notebooks on ocean.pangeo.io
- Store sonar data into a cloud-friendly format

Analysis Support

- Augment *echopype* with dimensionality reduction and pattern extraction techniques

Community Support

https://github.com/OSOceanAcoustics/echopype

Open-Source Ocean Acoustics
Home for open source tools and resources in ocean acoustics

echopype Contributors:

- Sven Gastauer (UCSD)
- Marian Peña (IEO Spain)
- Mark Langhirt (PSU)
- Erin LaBrecque (freelance)
- Emma Ozanich (UCSD)
- Aaron Marburg (APL-UW)

pyEcholab Developers:

- Zac Berkowitz (SOI)
- Rick Towler (AFSC)
- Chuck Anderson (NCEI)
- Veronica Martinez (NCEI)
- Pamme Crandall (NCEI)
- Carrie Wall (NCEI)

▪ Join us!

Links!

- *echopype* repository: https://github.com/OSOceanAcoustics/echopype
- *echopype* documentation: https://echopype.readthedocs.io/
- Sonar-netCDF Convention, v. 1.0, ICES CRR No. 341

Other Resources

- pyEcholab: https://github.com/CI-CMG/pyEcholab
- ESP3: https://bitbucket.org/echoanalysis/esp3/overview
- LSSS (Large Scale Survey System): https://cmr.no/projects/10396/lsss/
- Echogram: https://cran.r-project.org/package=echogram
- EchoView: https://www.echoview.com/
- etc. see echopype documentation

Acknowledgements

We thank all our sponsors for their generous funding.

NSF, Gordon and Betty Moore Foundation, Alfred P. Sloan Foundation, Washington Research Foundation