Interoperating ocean sonar data of heterogeneous sources using **echopype**

**Wu-Jung Lee** | Applied Physics Laboratory  
**Valentina Staneva** | eScience Institute  
**Kavin Nguyen** | Department of Physics

@leewujung

PICES 2019, Victoria, BC, Canada  
October 17, 2019
The acoustic data deluge

- We have become very good at collecting data
- Continuous data collection on numerous platforms
- Excellent opportunities to study marine ecosystems at scales never before possible
These new data bring many new challenges

It used to be...

Data from different instruments
- days, weeks

But now...

- Significantly increased data volume
- Heterogenous instrument sources
Challenges of the **Big Acoustic Data**

- **Interoperability**
  - Many manufacturers, many sonar models
  - Proprietary software or open-source software written in proprietary languages (e.g., Matlab)
Challenges of the Big Acoustic Data

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

A typical fisheries sonar data analysis pipeline

Download / access all raw data file → Pre-processing: calibration, denoising, etc. → Expert scrutiny → Advanced processing: classification, biomass estimation, etc. → Ecological & fisheries interpretation
Challenges of the **Big Acoustic Data**

- Interoperability
- Scalability
- Reproducibility
  - Currently mostly GUI-based: good for exploration but hard to reproduce
- Community
  - Workflow transformation
  - Community-driven development
  - Data convention
Vision

- Integrate data from heterogenous instrument sources
- Perform synoptic analyses over large spatial and temporal scales

- Long-term goal: make ocean sonar data an integrated component of standard oceanographic data sets

**Analysis methods**
- Interoperability
- Scalability
- Reproducibility
- Community

**Software tools**
Vision

- Interoperability
- Scalability
- Reproducibility
- Community

Software tools

- Integrate data from heterogenous instrument sources
- Perform synoptic analyses over large spatial and temporal scales

Long-term goal: make ocean sonar data an integrated component of standard oceanographic data sets
Current common workflow requires extensive data wrangling

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

Clean & combine data

Perform actual analysis

Ecological & fisheries interpretation
We can do better

- Leverage the scientific Python ecosystem!
Current common workflow requires extensive data wrangling.

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

Let’s build a package to handle these.

- Clean & combine data.
- Perform actual analysis.
- Ecological & fisheries interpretation.

SIMRAD
KONGSBERG
BioSor
ASL Environmental Sciences
EdgeTech
NORTEK
Our proposed workflow

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

Existing analyses

- echopyte
- scikit-learn
- TensorFlow
- PyTorch
- DASK
- xarray
- netCDF

Ecological & fisheries interpretation
echopype at a glance

- Follow ICES SONAR-netCDF4 for raw data storage when possible
- Take advantage of existing libraries
- Uniform interface to facilitate use and further development
- Calibration
- Other pre-processing

Visualization  Advanced analyses
Data representation

- Raw sonar data are just arrays

NumPy arrays
- Simple for Matlab users
- In-memory operations
- No labeled indexing for time, depth, frequency, etc.
- No metadata support

NetCDF4 gridded arrays
- More complex structure
- Allows direct access for efficient data selection/subsetting
- Labeled support for scientific interpretation
- Libraries for scalability and visualization
- Conversion tools for efficient cloud storage
Example notebook

- Watching a solar eclipse using an OOI sonar
- OOI = Ocean Observatories Initiative

The 2017 solar eclipse path

Access OOI data ➔ Process using echopype ➔ Combine and visualize

Access solar radiation data from US National Data Buoy Center (NDBC)

Try it out: 

Going forward 1: seek community input for sonar data convention

- echopype follows the SONAR-netCDF4 convention (ICES, May 2018) for raw data storage when possible
- Need convention for processed data to provide computational capability beyond raw data storage
- Generalizability vs efficiency?
- Yet another standard?
Going forward 2: grow echopype

- Include more sonar data formats, currently echopype supports:
  - Simrad EK60 .raw files
  - ASL AZFP .01A files
- Add advanced analysis and visualization routines
- Streamline cloud deployment
- Engage community in testing and development
THANK YOU!

- https://github.com/OSOceanAcoustics/echopype

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

- Join us!

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

AZFP Matlab toolbox developer
- Dave Billenness (ASL Env Sci)

pyEcholab developers
- Zac Berkowitz (SOI)
- Rick Towler (AFSC)
- Chuck Anderson, Veronica Martinez, Carrie Wall (NCEI)