BEeS
Heraklion
30 June - 3 July 2025

LifeWatch ERIC

hcmr ΕΛΚΕΘΕ

iMBBC ΙΘΑΒΒΥΚ

Session: Tracking the Wild: Unlocking Insights into Animal Movement, Behaviour, and Biologging

2 July 2025 | 14:30-16:30

Pangeo-fish: Scalable Geolocation Modelling of Fish Movements Using Biologging and Earth Science Data in the Pangeo Ecosystem

Etienne Cap[1], Quentin Mazouni[2], Justus Magin[1], Anne Fouilloux[2], Mathieu Woillez[3], Jean-Marc Delouis[1], Tina Odaka[1]

1 LOPS (Laboratory for Ocean Physics and Satellite remote sensing), UMR 6523, Univ Brest-Ifremer-CNRS-IRD, Plouzané, France
2 Simula Research Laboratory, Oslo, Norway
3 DECOD (Ecosystem Dynamics and Sustainability), IFREMER-Institut Agro-INRAE, Plouzané, France

Why studying fish movement?

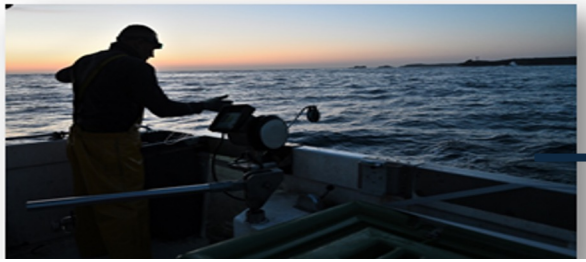
Motivation:

- **Food security and sustainable exploitation of natural resource**
- Study how fish movements and migrations shape the internal dynamics of populations
- knowledge, crucial for improving fishery management and define conservation area
- The International Council for the Exploration of the Sea (ICES):

Lack of information regarding fish spatial structure and essential fish habitats.

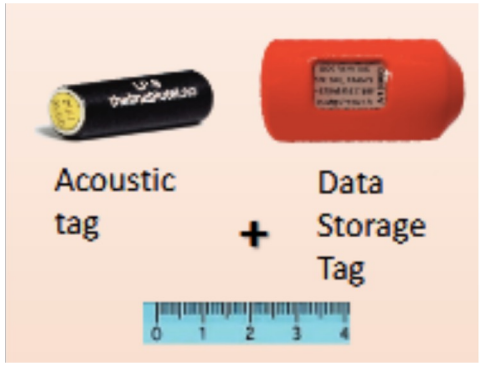
Bio-logging: studying fish movement

Fish capture

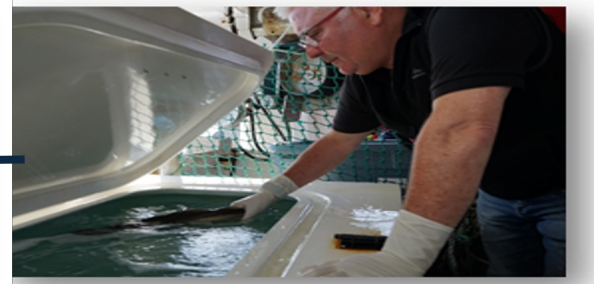


Stabulation, then anesthetize the fish

Implant the acoustic tag + Data Storage Tag (DST)



Fish release



Waking and housing fish

Data Storage Tag: Store time series of observed temperature, depth in a tag.
Acoustic Tag: tag emits signal, detect tagged fish using acoustic telemetry network
Combine both approaches together with earth science data: *Infer fish trajectory from individual environmental data histories using geolocation models*

Collecting data storage tag

The incentive program: Mobilizing citizens in the recovery of tagged fish

Project		Year	Nb tagged	Nb returned	%
PNMI		2010-2012	246	40	16.2
BARGIP		2014-2016	1220	482	39.5
FISH INTEL	Sea bass	2022	232	59	25.4
	Pollack	2022	70	19	27.1

Observation Frequency 90 seconds (most of it)

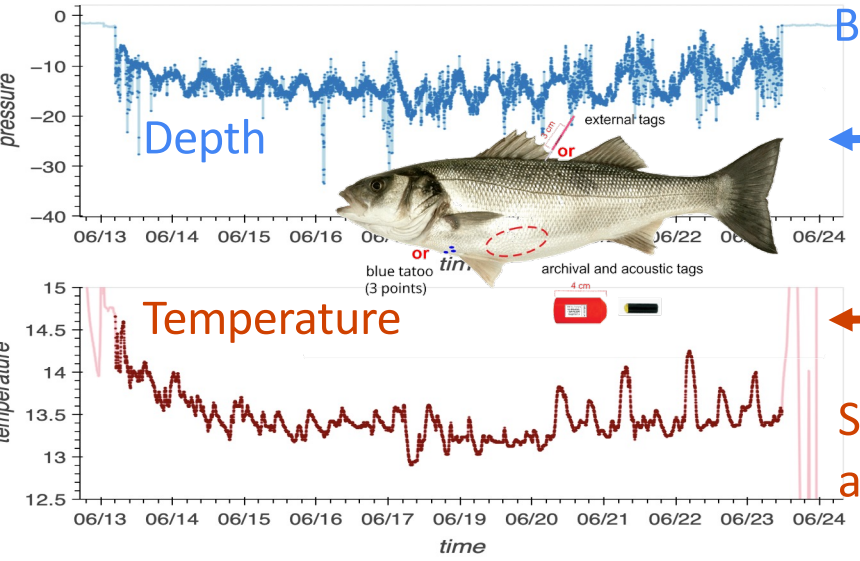
Acoustic tag detection: Acoustic telemetry network

- Stations' locations were based on **fishers' knowledge**
- **198 stations** have been deployed (October 2021 – October 2023)
- **Thelma Biotel** (TBR 800 RELEASE) in France, **VEMCO** in the UK and Belgium

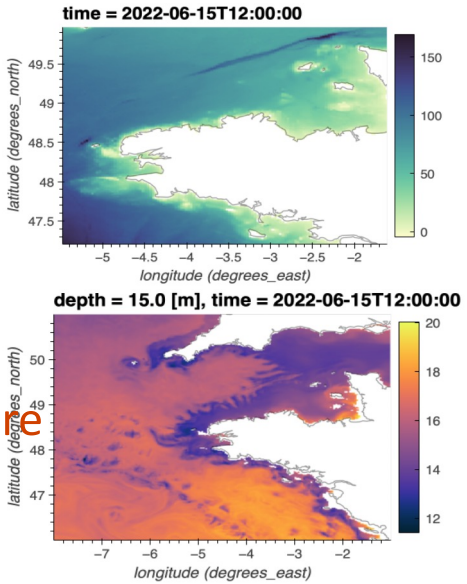
- The acoustic tag regularly **emits a unique identifier** over an extended period (1-5 years)
- The receiver can **detect acoustic tags** (in the tagged fish) that are **in the nearby area**
- **Data shared in ETN (European Tracking Network)**

Gonse et al. (2024)

Reconstruct Fish Track



Bio-logging data



Earth science data

Likelihood of Observed Temperature at depth by fish from earth science data

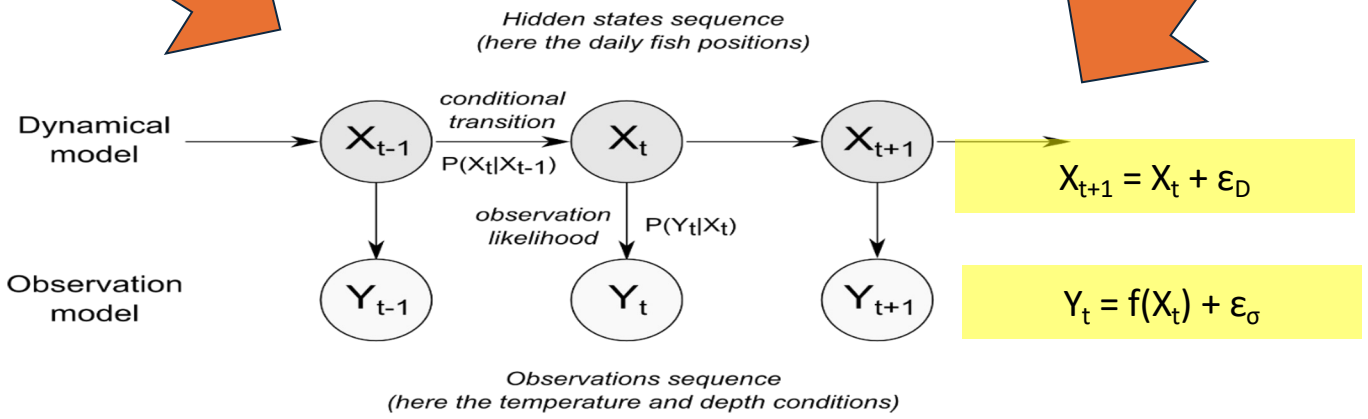
Constrains:

- ***acoustic detection,***
- ***bathymetry***
- ***release location***
- ***recapture location***

Model: Hidden Markov Method

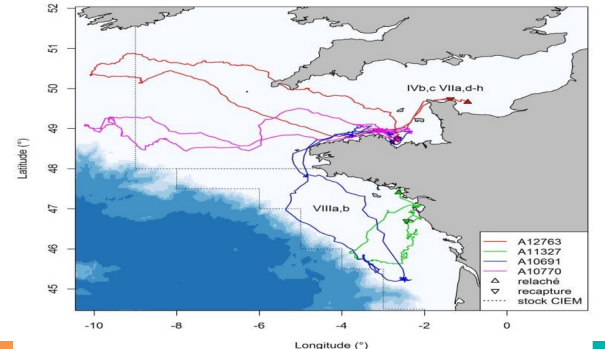
Geolocation model developed by Woillez et al. (2016), temperature and bathymetry as reference fields, adapted from Pedersen et al. (**2008**)

Hidden Markov Method



$X_{t+1} = X_t + \varepsilon_D$

$Y_t = f(X_t) + \varepsilon_\sigma$



Software for track reconstruction

- Quality and resolution (spatial and temporal) of earth science data (sea temperature) is crucial for track reconstruction.
- Original HMM software developed in 2016 **was not scalable.**
- Software workflow
  - Accessing in-situ observe environment data
  - Accessing big earth science data
  - Computation in python
  - Visualisation
- Fish track reconstruction: simillaire workflow as Big Data Geoscience!
- Let's **not reinventing the wheel**, but **using the already existing Scalable Big Data geoscience software stack**, such as PANGEO!

  => **Pangeo-fish**

PANGEO?

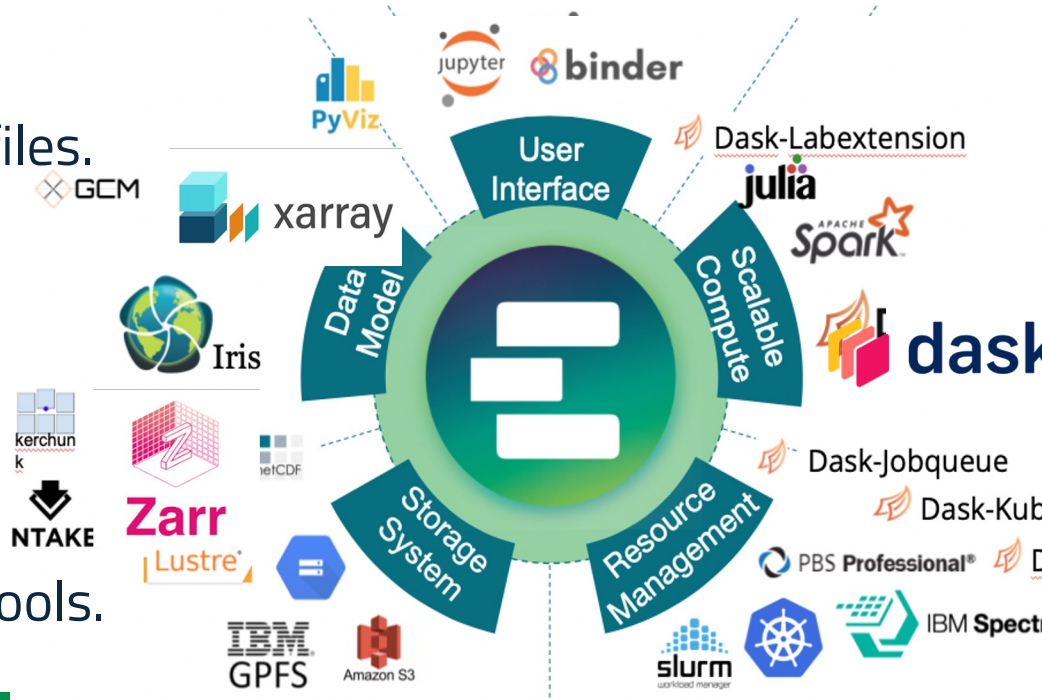
Pangeo is an open-source platform and ecosystem, designed for **scalable** and **reproducible data analysis in the Earth and climate sciences**. It brings together tools for **big data processing, cloud computing, and interactive scientific computing.**

Key Components:

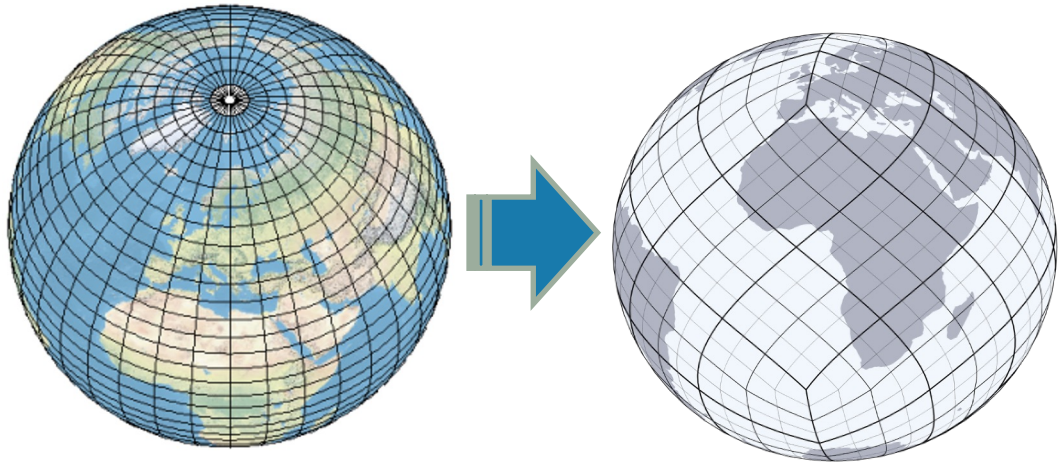
- **Xarray**: for working with labeled multi-dimensional arrays (like NetCDF data).
- **Dask**: for parallel computing on large datasets, enabling scalability from laptops to HPC and cloud.
- **Jupyter**: for interactive development and visualization.
- **Intake**: for managing and loading data catalogs.
- **Zarr**: for cloud-optimized data storage.
- **Kerchunk**: for virtualizing access to remote NetCDF/HDF5 files.

What Pangeo enables:

- Scalable workflows with terabytes of data
- ARCO (Analysis Ready Cloud Optimised) Format
- Analysis directly in the laptop / cloud / HPC environments.
- Collaborative and reproducible science using open-source tools.



Pangeo-fish and HEALPix



HEALPix is a type of DGGS

"A Discrete Global Grid System (DGGS) is a spatial reference system that uses a **hierarchical** tessellation of cells to partition and address the globe." OGC Abstract Specification, 2017

Open Geospatial Consortium

Equal Area, Seamless Global Coverage, and Multi-Scale.

Expand possible fish tracking system to globa applicationl!!

```
[20]: test_plot.dggs.explore(center=0, cmap="ocean", alpha=0.5)
```

[20]:

xdggs

PANGEO

Data Consistency with HEALPix(DGGS)

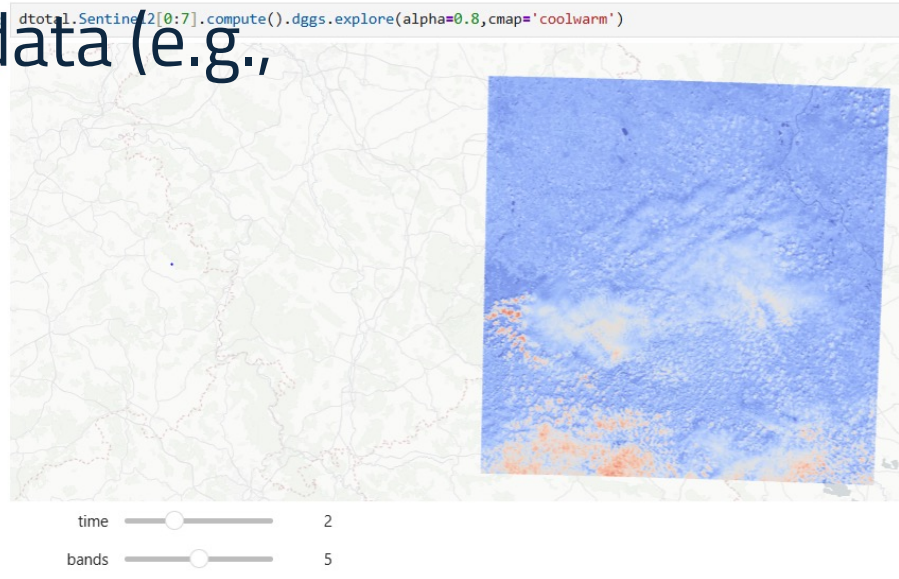
- Overcoming challenges related to **unequal cell surface areas**.
- **xdggs** to easily access and visualise HEALPix data

Pangeo ecosystem and HEALPix for computing fish-track = Pangeo-fish

GRID4EARTH: HEALPix

Common DGGS for DestinE and Copernicus Sentinel Data

- Part of ESA's **Digital Twin Earth (DTE) Framework**
- **GRID4EARTH** project aims to develop a unified, standardized data infrastructure
- Based on a **Discrete Global Grid Systems (DGGS), HEALPix**
- Enables seamless **integration, visualization, and analysis** of Earth Observation (EO) data
- Addresses growing **complexity and volume** of EO data (e.g., **Copernicus** and **DestinE** programs)

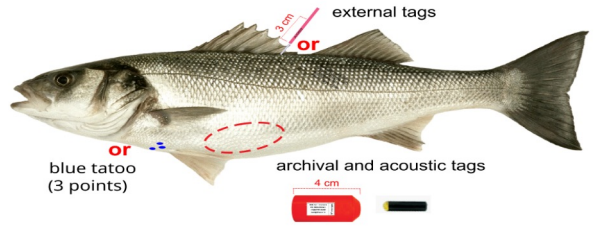


Funded by:
The European Space Agency

Consortium Members:

IFREMER, CNRS, Simula, The University of Tartu, GEORODE

Conclusion



- Pangeo-Fish is a **generic and reusable tool** for analyzing fish biologging data.
- Promotes **synergy between biology and Earth science** through the use of **shared, open-source tools.**
- Thanks to Pangeo eco system and HEALPix, the geolocation model is ready for intensive use with higher-resolution geophysical reference data, enabling more accurate reconstructions.
- Submitted to JOSS(under review)
- 📦 Install via **pip install pangeo-fish**
- 🔗 GitHub: github.com/pangeo-fish/pangeo-fish
- This work demonstrates how open, FAIR, and scalable eScience tools can bridge data gaps in marine ecology and support biodiversity conservation in a changing ocean.

Thank you for your attention!

Questions?
Tina.odaka@ifremer.fr

BEeS
The LifeWatch ERIC Biodiversity & Ecosystem eScience Conference

Heraklion, 30 June - 3 July 2025