

pangeo-fish: A Python package for studying fish movement using biologging and earth science data

Justus Magin¹, Quentin Mazouni², Etienne Cap¹, Marine Gonse³,
Mathieu Woillez³, Anne Fouilloux², Jean-Marc Delouis¹, and Tina
Odaka¹

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

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Richard Littauer

Reviewers:

- @platipodium
- @rsignell
- @maxrjones

Submitted: 07 March 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Geo-referenced data plays an important role in understanding and conserving natural resources, particularly when investigating biological phenomena such as fish migration and habitat uses. Biologging, the practice of attaching small devices (called *tags*) to animals for recording behavior, physiology, and environmental data, proves to be invaluable in this field.

As fish can not be tracked directly using tracking devices such as GPS receivers, geolocation models have emerged to estimate fish positions by correlating individual series of physical measurements — e.g. temperature and pressure records — with geophysical reference fields — oceanic temperature and bathymetry — derived from satellite observations and hydrodynamical model outputs.

Beside the difficulty of working with vast earth science datasets (due to their size and diversity), there is no open source implementation for biologged fish tracking. Yet, these fish geolocation models are critical for better understanding fish behavior and are nowadays seen as a powerful tool by policy makers to improve fish management and conservation.

To address this challenge, we developed a Python package, named **pangeo-fish**, for fish tracking estimation. It is based on the **Pangeo** ecosystem, which offers a unique interoperable, scalable, open source environment for interactive data analysis in the field of big data marine and geoscience.

Statement of need

Biologging consists in attaching onto (or sometimes inserting into) an animal an electronic device that will record in its memory physical and/or geochemical parameters as a function of time so that scientists can reconstruct the activity of the animal, the characteristics of the environment it travels in and the interactions between the two.

These tools can provide a wealth of information on the behaviors and movements of free-swimming marine animals, such as diving and activity patterns, energy use and interaction with environment.



Figure 1: Promotion of the FISH-INTEL tagging campaign.

35 However, unlike terrestrial animals or marine mammals, whose positions can be directly
36 estimated using ARGOS or GPS technologies, tracking fish underwater is challenging. To
37 address this issue, various tagging experiments have been conducted on a variety of fish species
38 (Carla et al., 2021; Mucientes et al., 2021), and methods have been proposed for approximating
39 the fish locations, referred to as geolocation models (de Pontual et al., 2023; Woillez et al.,
40 2016).

41 For studying fish movements, the two widely used electronic tagging technologies are acoustic
42 telemetry and Data Storage Tags (DST, or archival tags). Acoustic telemetry involves a tag
43 that emits an acoustic signal containing a unique ID and possibly sensor data. This signal can
44 be detected by an acoustic receiver when the tagged animal is within range, and the detection
45 data is retrieved from the receiver. As such, acoustic tags do not need to be recovered, but
46 there is no guarantee that the tagged fish will swim around the receivers network.

47 In contrast, archival tags store sensor measurements at set intervals in their memory. To access
48 the logged data, these tags must either be recovered (which mostly depends on fishers and
49 the local population living along the coast) or transmit their information via satellite. In the
50 former case, tagging campaigns usually promote and possibly reward tag or fish captures (see
51 for instance, the advertisement from the FISH-INTEL campaign on Figure 1). The data from
52 archival tags can offer detailed insights into vertical movement patterns (Heerah et al., 2017)
53 or environmental preferences (Carla et al., 2021; Righton et al., 2010), and can be used to
54 reconstruct migration paths through geolocation modeling.

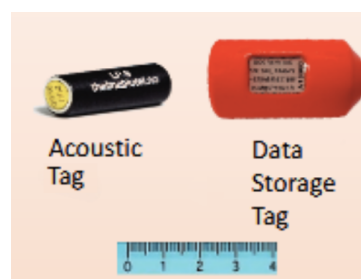


Figure 2: Example of an acoustic tag (on the left) and a DST (on the right). See the centimeter scale for size reference.

55 Figure 2 shows an example of an acoustic tag as well as a DST.

The estimation of fish positions depends on the likelihood of the observed data from the DTS's logs, such as temperature at specific depths, alongside the reference geoscience data such as satellite observations and ocean dynamic models. Some approaches can enhance the accuracy of the model's predictions by using additional information, such as telemetric detection data from the acoustic tags mentioned above (Goossens et al., 2023). The use of oceanic models with high spatial and temporal resolutions can significantly improve the accuracy of reconstructed fish tracks. However, higher resolutions involves more data, that requires significant computing power and storage capacity. The Pangeo community handles these challenges, by fostering an ecosystem of interoperable, scalable, open source tools for interactive data analysis in the field of big data marine and geoscience. Therefore, the Pangeo ecosystem represents a powerful mean through which biologists can analyze more easily their biologging data and improve fish geolocation modelling. Not only their results would eventually guide policy makers to manage fish stock in a more sustainable way, they could also be used to forecast potential movement changes due to the ongoing climate change.

Unfortunately, the research community lacks of adaptable, scale and open source implementations of geolocation models. **pangeo-fish** is a Python package that fills this gap.

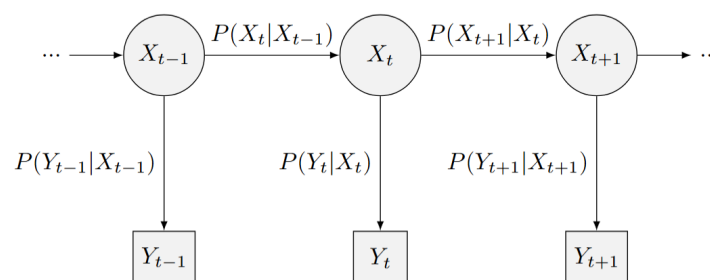
As its name suggest, the software has been designed to be used within the Pangeo ecosystem on several aspects, therefore accounting for both the users' needs (user-friendly API and meaningful result visualization) and the computational challenges. In particular, **pangeo-fish** has a robust data model based on Xarray and scales computation with Dask.

Data loading processes are furthermore streamlined by libraries like intake, kerchunk or fsspec, and the previously mentioned xarray data model enables interactive visualization of the results thanks to tools such as the hvplot library and the JupyterLab environment.

Similarly, pangeo-fish's I/O operations are automatically distributed with the combination of Dask and Zarr. The Pangeo software stack provides researchers with the necessary tools to access reference data and perform intensive computations in a scalable and interactive manner. pangeo-fish gives ecologists a user-friendly tool for inferring fish locations from archival tag data, hence filling the gap between their expertise and the Pangeo's environment capabilities.

Geolocation Model

pangeo-fish implements a method well established in the fish trajectory reconstruction literature Goossens et al. (2023). It consists of a Hidden Markov Model (HMM).



- $P(X_t|X_{t-1})$: Transition Probabilities
- $P(Y_t|X_t)$: Emission Probabilities

Figure 3: Illustration of the Hidden Markov Model. The hidden states X_t describe the fish's positions, and the emission probabilities $P(Y_t|X_t)$ correspond to the likelihood of observing the fish at time t .

As illustrated in Figure 3, the latent (or *hidden*) states X_t of the HMM infer the (daily or hourly) fish's positions, and the observation process relates the sensor records with the oceanic data. The transition matrix between the hidden states is modelled by a Brownian motion parametrized by σ . As such, fitting the geolocation model for a tag's records aims to determine the value of σ that maximizes the likelihood of the state sequence (i.e., the fish's trajectory) given the observations. The optimal likelihood value reflects the level of residual inconsistency between the tag observed (recorded) and reference data.

Conclusion

pangeo-fish is a Python package that implements a geolocation model, based on a Hidden Markov Model, for estimating fish positions from archival tag and oceanic data. Designed to work with the Pangeo ecosystem, it aims to support the ecologists with their research, by handling backend processes — such as data loading or parallel computation — while exposing a user-friendly interface to manage their archival tag data and run the geolocation model.

Acknowledgements

- T Odaka, JM Delouis and J Magin are supported by the CNES Appel, a projet R&T R-S23/DU-0002-025-01.
- T Odaka, JM Delouis and M Woillez are supported by the TAOS project funded by the IFREMER via the AMII OCEAN 2100 programme.
- Q Mazouni, M Woillez, A Fouilloux and T Odaka are supported by Global Fish Tracking System (GFTS), a Destination Earth use case procured and funded by ESA (SA Contract No. 4000140320/23/I-NS, DESTINATION EARTH USE CASES - DESP USE CASES - ROUND (YEAR 2023))
- M Woillez and T Odaka are supported by Digital Twin of the Ocean - Animal Tracking (DTO Track), a European project that aims to create a digital twin of the North Sea using animal tracking data and funded by the Sustainable Blue Economy Partnership (ANR-24-SBEP-0001)

References

- Carla, F., Villegas-Ríos, D., Moland, E., & Olsen, E. (2021). Sea temperature effects on depth use and habitat selection in a marine fish community. *Journal of Animal Ecology*, 90. <https://doi.org/10.1111/1365-2656.13497>
- de Pontual, H., Heerah, K., Goossens, J., Garren, F., Martin, S., Le Ru, L., Le Roy, D., & Woillez, M. (2023). Seasonal migration, site fidelity, and population structure of European seabass (*Dicentrarchus labrax*). *ICES Journal of Marine Science*, 80(6), 1606–1618. <https://doi.org/10.1093/icesjms/fsad087>
- Goossens, J., Woillez, M., LeBris, A., Verhelst, P., Moens, T., Torreele, E., & Reubens, J. (2023). Acoustic and archival technologies join forces: A combination tag. *Methods in Ecology and Evolution*, 14(3), 860–866. <https://doi.org/10.1111/2041-210X.14045>
- Heerah, K., Woillez, M., Fablet, R., Garren, F., Martin, S., & De Pontual, H. (2017). Coupling spectral analysis and hidden markov models for the segmentation of behavioural patterns. *Movement Ecology*, 5, 1–15. <https://doi.org/10.1186/s40462-017-0111-3>
- Mucientes, G., Leeb, K., Straßer, F.-E., Villegas-Ríos, D., & Alonso-Fernández, A. (2021). Short-term survival, space use and diel patterns of coastal fish species revealed from 'solo datasets'. *Marine and Freshwater Behaviour and Physiology*, 54. <https://doi.org/10.1080/10236244.2021.1912604>

- 131 Pedersen, M. W., Righton, D., Thygesen, U. H., Andersen, K. H., & Madsen, H. (2008).
132 Geolocation of north sea cod (*gadus morhua*) using hidden markov models and behavioural
133 switching. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(11), 2367–2377.
134 <https://doi.org/10.1038/42622>
- 135 Righton, D. A., Andersen, K. H., Neat, F., Thorsteinsson, V., Steingrund, P., Svedäng, H.,
136 Michalsen, K., Hinrichsen, H.-H., Bendall, V., Neuenfeldt, S., & others. (2010). Thermal
137 niche of atlantic cod *gadus morhua*: Limits, tolerance and optima. *Marine Ecology Progress*
138 *Series*, 420, 1–13. <https://doi.org/10.3354/meps08889>
- 139 Woillez, M., Fablet, R., Ngo, T.-T., Lalire, M., Lazure, P., & de Pontual, H. (2016). A
140 HMM-based model to geolocate pelagic fish from high-resolution individual temperature
141 and depth histories: European sea bass as a case study. *Ecological Modelling*, 321, 10–22.
142 <https://doi.org/10.1016/j.ecolmodel.2015.10.024>

DRAFT