The Journal of Open Source Software

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

Justus Magin [1], **Quentin Mazouni** [2], **Etienne Cap** [1], **Marine Gonse** [3], **Mathieu Woillez** [3], **Anne Fouilloux** [4], **Jean-Marc Delouis** [1], and **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 **4** LifeWatch ERIC, Seville, Spain

DOI: 10.21105/joss.08332

Software

- Review
- Repository
- Archive

Editor: Richard Littauer

Reviewers:

- @platipodium
- @rsignell

Submitted: 07 March 2025
Published: 03 June 2026

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, is 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.

Besides 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 understanding fish behavior better and are currently 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 fields of marine big data and geoscience.

Statement of need

Biologging consists of attaching an electronic device onto (and sometimes inserting into) an animal, which 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 environmental interactions.

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

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

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

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

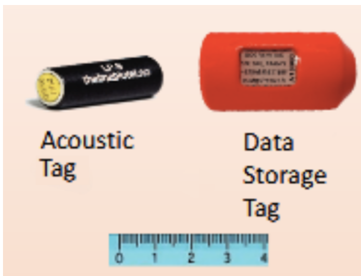

Figure 2: Example of an acoustic tag (on the left) and a DST (on the right). See the centimeter scale for size reference.

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 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 telemetry 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 involve more data, which 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 fields of marine big data and geoscience. The Pangeo ecosystem enables biologists to more easily analyze their biologging data and improve fish geolocation modelling. Their results can guide policy makers to sustainably manage fish stocks, and can also be used to forecast potential long-term movement changes due to climate change.

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

As its name suggests, the software has been designed to be used within the Pangeo ecosystem on several aspects, accounting for both the users' needs (through a user-friendly API and meaningful result visualization) and 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`. `Xarray` enables interactive visualization of the results using 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.

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).

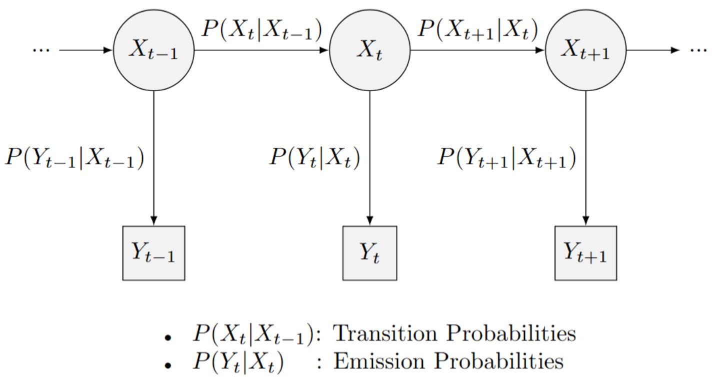

Figure 3: Illustration of the Hidden Markov Model. The hidden states Xt describe the fish's positions, and the emission probabilities $P(Y_t|X_t)$ correspond to the likelihood of observing the fish at time t.

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 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 geolocation models.

Acknowledgements

T.O., J.M.D. and J.M. were supported by the CNES R&T project (R-S23/DU-0002-025-01). T.O., J.M.D. and M.W. received support from the TAOS project funded by IFREMER through the AMII OCEAN 2100 programme. Q.M., E.C, M.W., A.F. and T.O. were supported by the Global Fish Tracking System (GFTS) initiative funded by ESA (Contract No. 4000140320/23/I-NS, Destination Earth Use Cases 2023). Finally, M.W., E.C and T.O. acknowledge support from the Digital Twin of the Ocean – Animal Tracking (DTO Track) project 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

Pedersen, M. W., Righton, D., Thygesen, U. H., Andersen, K. H., & Madsen, H. (2008). Geolocation of North Sea cod (Gadus morhua) using hidden Markov models and behavioural switching. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(11), 2367–2377. https://doi.org/10.1038/42622

Righton, D. A., Andersen, K. H., Neat, F., Thorsteinsson, V., Steingrund, P., Svedäng, H., Michalsen, K., Hinrichsen, H.-H., Bendall, V., Neuenfeldt, S., & others. (2010). Thermal niche of Atlantic cod Gadus morhua: limits, tolerance and optima. *Marine Ecology Progress Series*, *420*, 1–13. https://doi.org/10.3354/meps08889

Woillez, M., Fablet, R., Ngo, T.-T., Lalire, M., Lazure, P., & de Pontual, H. (2016). A

HMM-based model to geolocate pelagic fish from high-resolution individual temperature and depth histories: European sea bass as a case study. *Ecological Modelling*, *321*, 10–22. https://doi.org/10.1016/j.ecolmodel.2015.10.024