Interpolation of Plankton Continuous Recorder data using a neural network technique

Charles Troupin[1] (ctroupin@uliege.be), Alexander Barth[1] (abarth@uliege.be), Pierre Hélaouët[2] (pihe@mba.ac.uk)

[1]University of Liège (Belgium)
[2]Marine Biological Association (UK)

Summary
We present the results obtained by the application of a neural network method, DINCAE, on *Calanus finmarchicus* and *Calanus helgolandicus* observations collected by the *Continuous Plankton Recorder*. The main features (spatial distribution, seasonal cycle) are well reproduced in the interpolated fields, even if the amplitude of the variation is dampened by the gridding method.

Data and methods
Continuous Plankton Recorders (CPR) are high-speed plankton samplers towed by ships of opportunity. The same sampling method has been used for decades, making the dataset particularly relevant to study long-term changes.
We focus on *Calanus finmarchicus* and *Calanus helgolandicus*, which constitute key contributors to the zooplanktonic ecosystem. They are both known to be influenced by environmental parameters, such as the bathymetry, the surface sea water temperature [Hélaouët and Beaugrand, 2007].
The uneven sampling effort is highlighted in Figure 1, where the main routes are clearly visible. The dataset contains a total of 250021 records (time, position and counts of different taxa and taxonomic groups).
The *Data-Interpolating Convolutional Auto-Encoder* (DINCAE) method [Barth et al., 2022] is applied to create gridded fields from the in-situ observations. Environmental data are also considered in the neural network. The code is written in Julia and available from https://github.com/gher-uliege/DINCAE.jl.
Several *hyperparameters* have to be set in order to run DINCAE. Some of them were set by hand, i.e., by using standard values, while the most relevant parameters (learning rate, number of epochs and Laplacian penalty coefficient) are selected using a cross-validation approach.

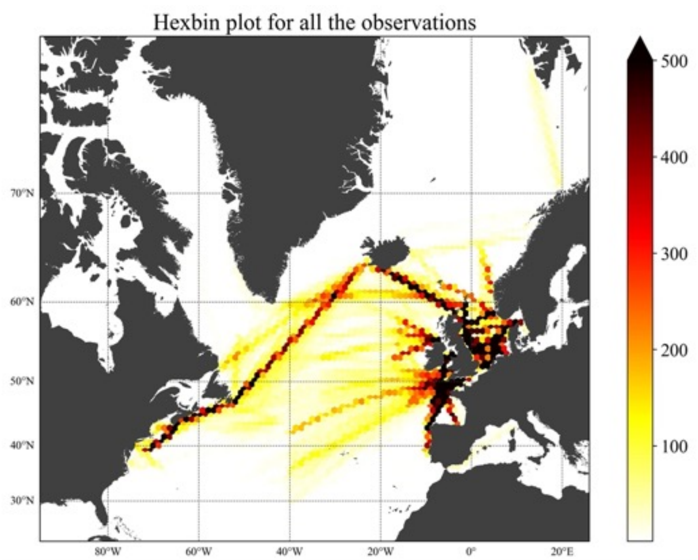

Figure 1 Spatial distribution of Calanus Finmarchicus and *Calanus Helgolandicus*.

Results and validation

Several experiments have been conducted with different configurations:

- input data preparations: applying a transformation (logarithm) or a quality control (removing values above the 95th percentile), considering different time periods (climatology, monthly fields, monthly fields with 2 years before/after the year of interest, …);
- environmental data: bathymetry, sea surface temperature, distance to closest coast.

The validation is performed by discarding 10% of the observations, performing the DINCAE analysis and then comparing the values of the gridded field at the locations of the discarded data with the discarded values.

The results display the main features (spatial distribution and seasonal cycle) once the hyperparameters have been tuned and when the bathymetry and the sea surface temperature are considered in the neural network.

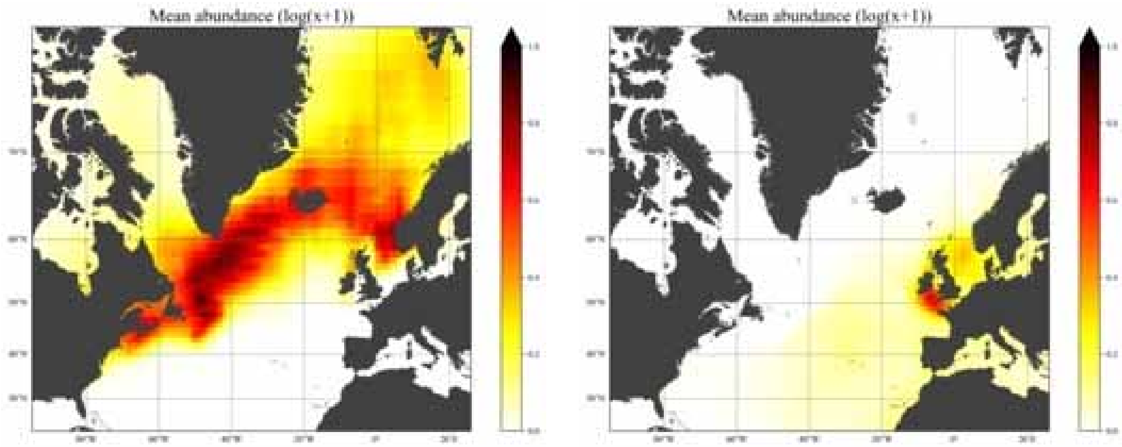

Figure 1 Mean fields (all years and months) of *Calanus finmarchicus* and *Calanus helgolandicus* produced with DINCAE.

Conclusions

This work presents a first attempt to interpolate in situ, zooplankton data using a neural network technique. Future work will focus on the following aspects:

1. Further optimisation of the hyperparameters, since here the attention was specifically on 3 of them: (learning rate, number of epochs and Laplacian penaly coefficient.
2. Interpolation of other taxa or other taxonomic groups, with a particular interest on the *Total Dinoflagellates* or the *Total Copepods*.
3. Use of other environmental variables such as the chlorophyll concentration and its horizontal gradient.

References

Barth A., Alvera-Azcárate A., Troupin C., & Beckers J.-M. (2022). *DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations*. Geoscientific Model Development, 15(5), 2183–2196. https://doi.org/10.5194/gmd-15-2183-2022

Helaouët P. and Beaugrand G. (2007). *Macroecology of Calanus finmarchicus and C. helgolandicus in the North Atlantic Ocean and adjacent seas*, Marine Ecology Progress Series, 345, 147–165, https://doi.org/10.3354/meps06775