# Reconstructing the Seasonal Cycle of Upper-Ocean Biogeochemical Profiles in the Norwegian Sea with BGC Argo–Informed Machine Learning

Fabio Mangini[1], Tsuyoshi Wakamatsu[1*], Roshin P. Raj[1], Juliano Ramanantsoa[2], Antonio Bonaduce[1], Julien Brajard[1], Laurent Bertino[1]

[1]Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research, Bergen, Norway

[2]Geophysical Institute, University of Bergen, Bergen, Norway

*corresponding author : tsuyoshi.wakamatsu@nersc.no

NERSC NANSEN ENVIRONMENTAL AND REMOTE SENSING CENTER THE NANSEN CENTER • BERGEN • NORWAY

## Introduction

We use a machine-learning approach, known as Profile Hidden Markov Model (PROFHMM; Charantonis et al., 2015), to reconstruct the seasonal cycle of chlorophyll-a (Chl-a) and particulate organic carbon (POC) in the Nordic Seas down to 400 m over the last 15 years circa. The method relies on surface variables and mixed-layer depth, since ocean properties are commonly better observed at the surface than at depth.

The poster present a technical evaluation of the performance of PROFHMM. The method has previously been used only with model outputs (Charantonis et al., 2015; Galmiche et al., 2019). In this study, we first apply it to observational data from satellites and BGC ARGO. Then, we assess the method's performance, using a case-study profile and metrics against individual profiles and the climatologies of the Norwegian and Lofoten Basins.

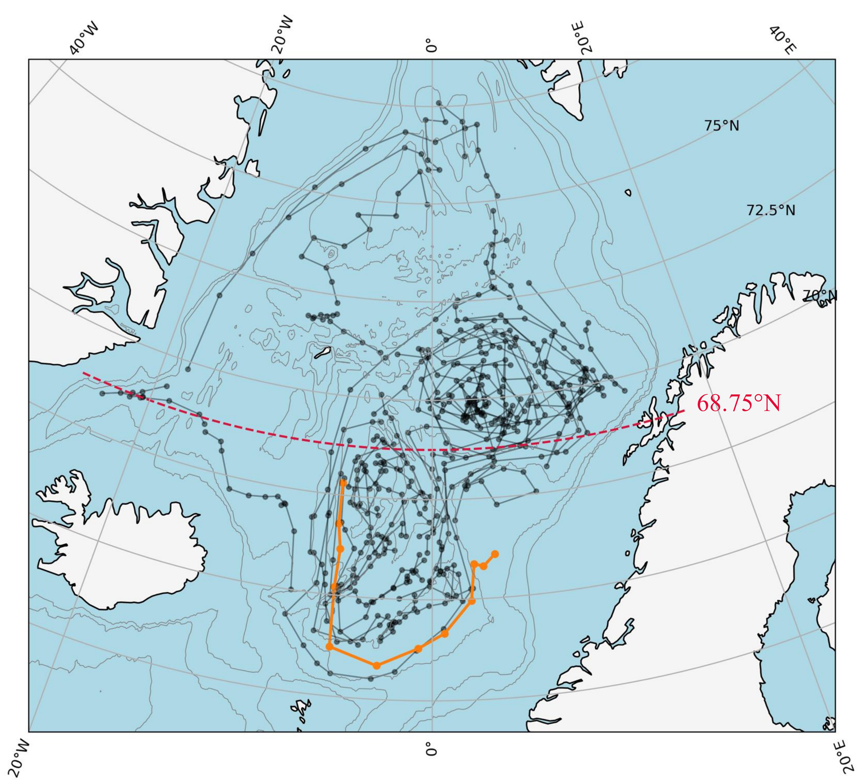
Despite the technical nature of the poster, we should highlight that the broader aim of this project is scientific. For instance, this approach can help investigate climate-driven changes in the biology of the Nordic Seas.

## Datasets

| Variable, [abbreviation; unit] | Source (dataset name) |
|---|---|
| Temperature [T; K] | Argo GDAC |
| Salinity [S; PSU] | Argo GDAC |
| Chlorophyll-a [Chl-a; mg/m$^3$] | Argo GDAC |
| Particulate Backscattering Coefficient 700nm [$bbp_{700}$; m$^{-1}$] | Argo GDAC |
| Sea-Surface Temperature [SST; K] | ESA CCI (SST_GLO_SST_L4_REP_OBSERVATIONS_010_024) |
| Sea-Surface Chlorophyll-a Concentation [Chl-a; $\log_{10}$(mg/m$^3$)] | Monthly gap filled OC-CCI chlorophyll-a (Ford et al., 2026) |
| Surface solar radiation downward clear-sky [ssrdc; J/m$^2$] | ERA5 (surface_solar_radiation_downward_clear_sky) |

## Datasets reprocessing

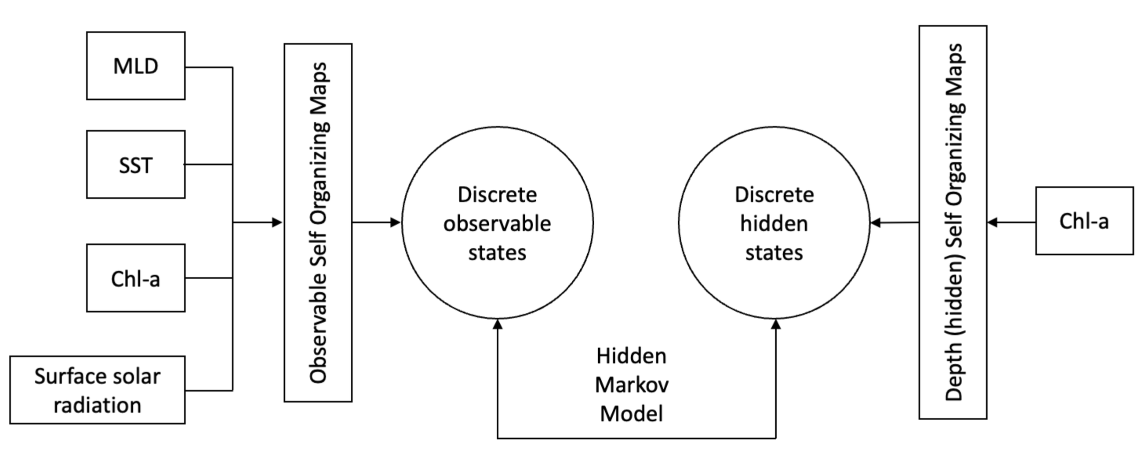
This study uses BGC ARGO floats within the Nordic Seas between 2011 and 2024. Temperature and salinity are used to estimate the mixed-layer depth (mld), an input to PROFHMM along with surface variables. Chl-a profiles are used both to train the model and, together with $bbp_{700}$, to estimate POC.




We split the dataset into annual segments and retain only complete one-year records, resulting in 43 monthly annual time series (shown in the map to the right).

Of the 43 segments, 33 are used for training and 10 for validation. Surface data were then collocated with the 43 BGC ARGO floats.

## Profile Hidden Markov Model (PROFHMM)

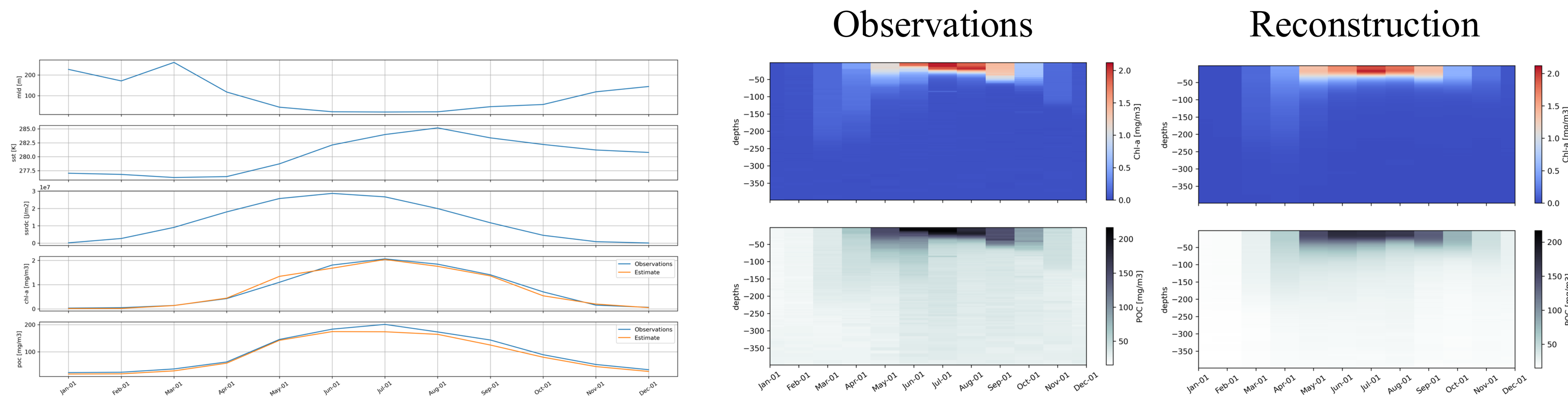



PROFHMM assumes that the value of the seasonal cycle at a given month depends only on the previous month, and uses a first-order Markov model. More precisely, it employs a hidden Markov model because it seeks to estimate subsurface (hidden) variables from surface (observable) variables. The connection between surface and subsurface is established using discretized fields generated with self-organizing maps (SOM).

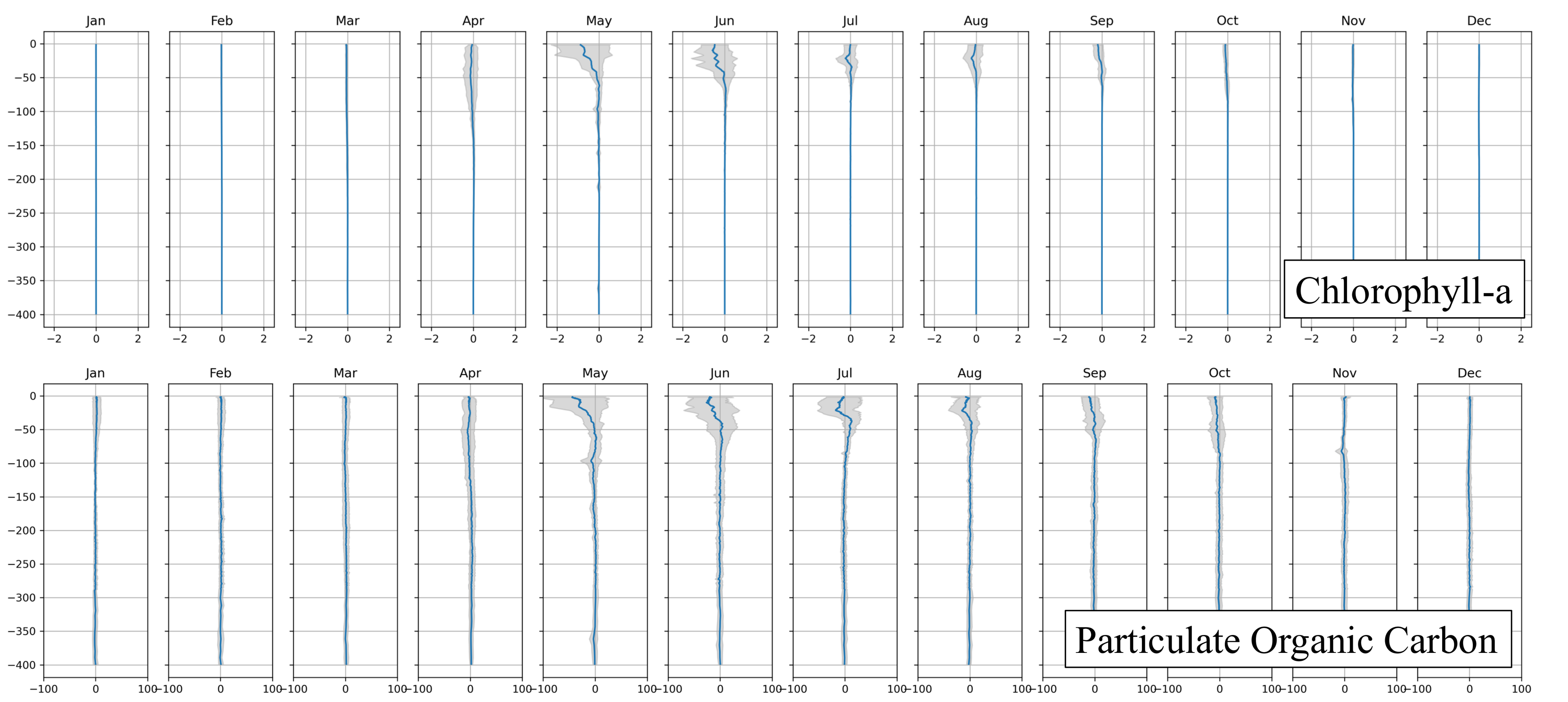
The model is run using the SubMAPP package (link to the tutorial in the References section). To ensure robustness, we adopt an ensemble approach: the PROFHMM is run 100 times, and the ensemble mean is then used as the final estimate.

## Results

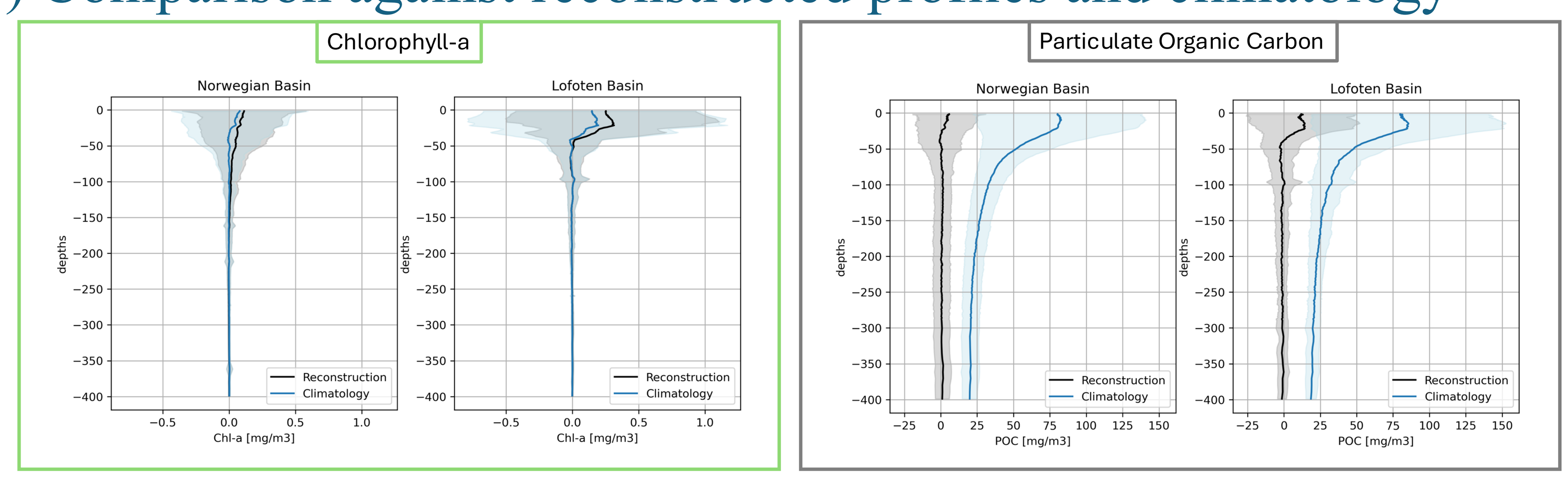
### 1) Example: ARGO ID “6903554” in 2020



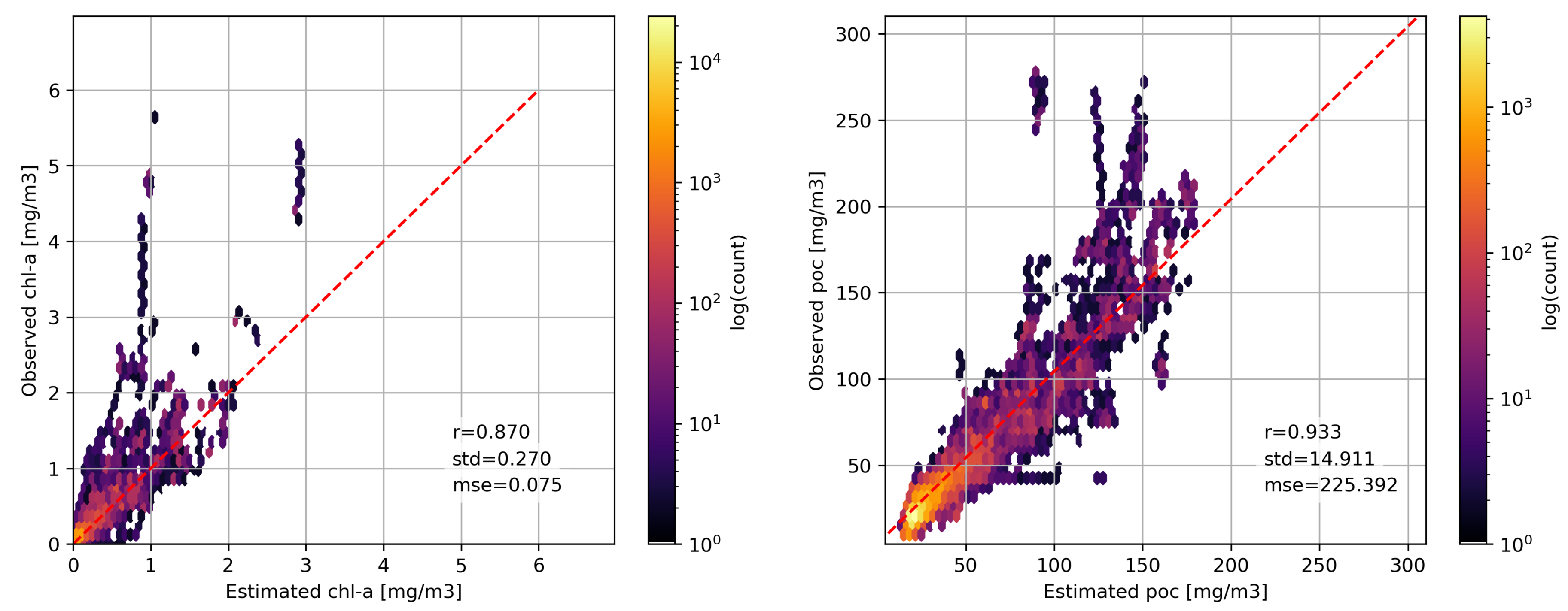

### 2) Profiles of bias and standard deviation




### 3) Comparison against reconstructed profiles and climatology




### 4) Comparison between observations and all estimated profiles




## Conclusions and future work

PROFHMM successfully reconstructs Chl-a and POC to 400 m, but the reconstructed amplitudes are lower than observations. Also, the climatologies of the Norwegian and the Lofoten Bains may be preferable to the reference profiles.

Future work might include using the full temporal resolution of the data, extending the method to 2D fields, and relaxing the requirement for complete annual time series.

## References

1. Charantonis, A. A., Badran, F., and Thiria, S. Retrieving the evolution of vertical profiles of Chlorophyll-a from satellite observations using Hidden Markov Models and Self-Organizing Topological Maps. *Remote Sensing of the Environment*, vol. 163, pp. 229-239, 2015.
2. Ford, D. J., Kulk, G., Sathyendranath, S., and Shutler, J. D.: Decadal and spatially complete global surface chlorophyll-a data record from satellite and BGC-Argo observations, *Earth Syst. Sci. Data*, 18, 569–584, https://doi.org/10.5194/essd-18-569-2026, 2026.
3. Galmiche, N., Brajard, J., Charantonis, A., Wakamatsu, T. Impact of sparse profile sampling on the reconstruction of sub-surface ocean temperature from surface information. *Proceeding of the 9th International Workshop on Climate Informatics*: CI 2019 (No. NCAR/TN-561+PROC). doi:10.5065/y82j-f154.
4. SubMAPP package tutorial available at: https://github.com/nansencenter/SubMAPP_tutorial.