The Synergy of Data From Profiling Floats, Machine Learning and Numerical Modeling: Case of the Black Sea Euphotic Zone

E. V. Stanev , K. Wahle , and J. Staneva

COSS-TT meeting
Montreal 2-4 May 2023

Stanev, E. V., Wahle, K., & Staneva, J. (2022). The synergy of data from profiling floats, machine learning and numerical modeling: Case of the Black Sea euphotic zone. Journal of Geophysical Research: Oceans, 127, e2021JC018012. https://doi. org/10.1029/2021JC018012

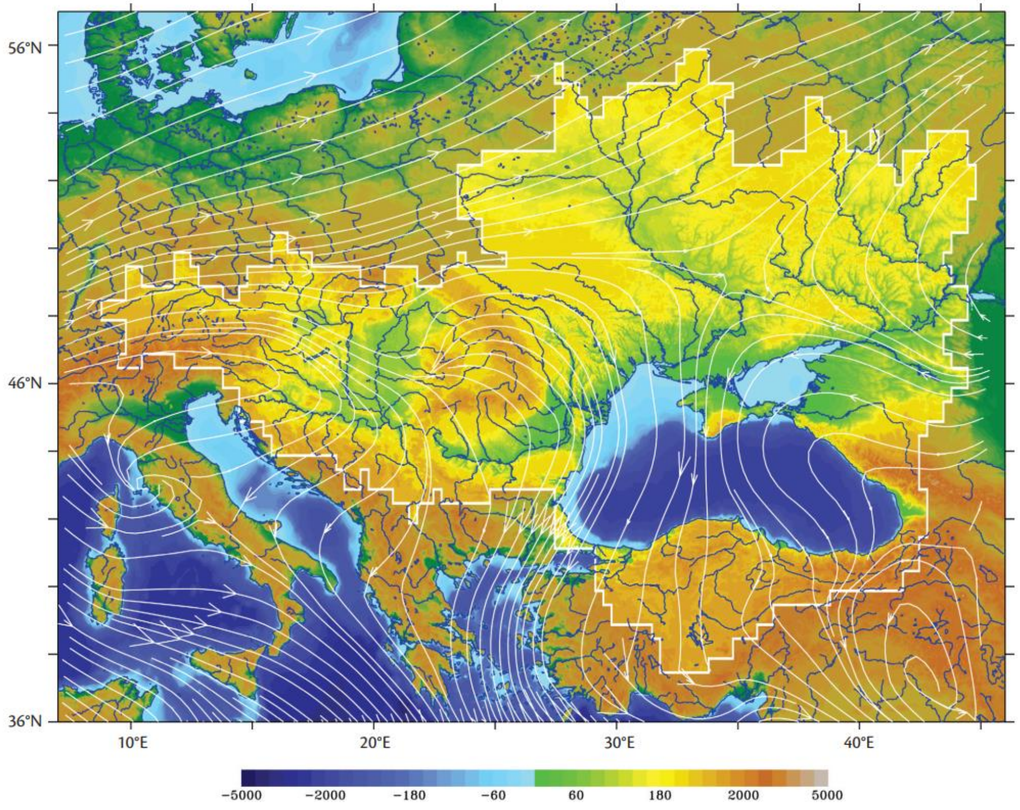
Recent developments

- *attributable to satellites*
- *development of sensors (BGC-Argo)*
- *advancement in numerical BGC modeling*

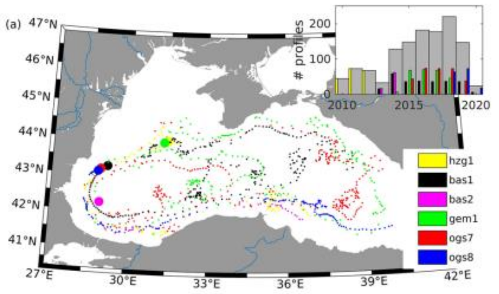
Do available observations represent adequately the 4D BGC dynamics?

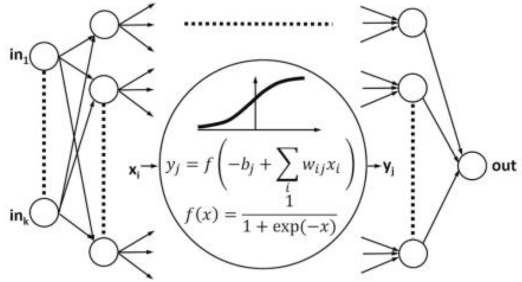
How can ML help?

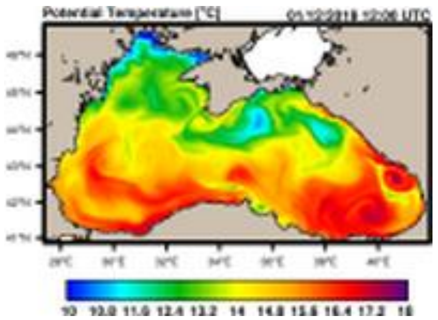
I will present an example of utilizing ML based on BGC-Argo data from the Black Sea region.
We investigate the photic layer.



BGC Argo + *AI* + *CEMS physics*



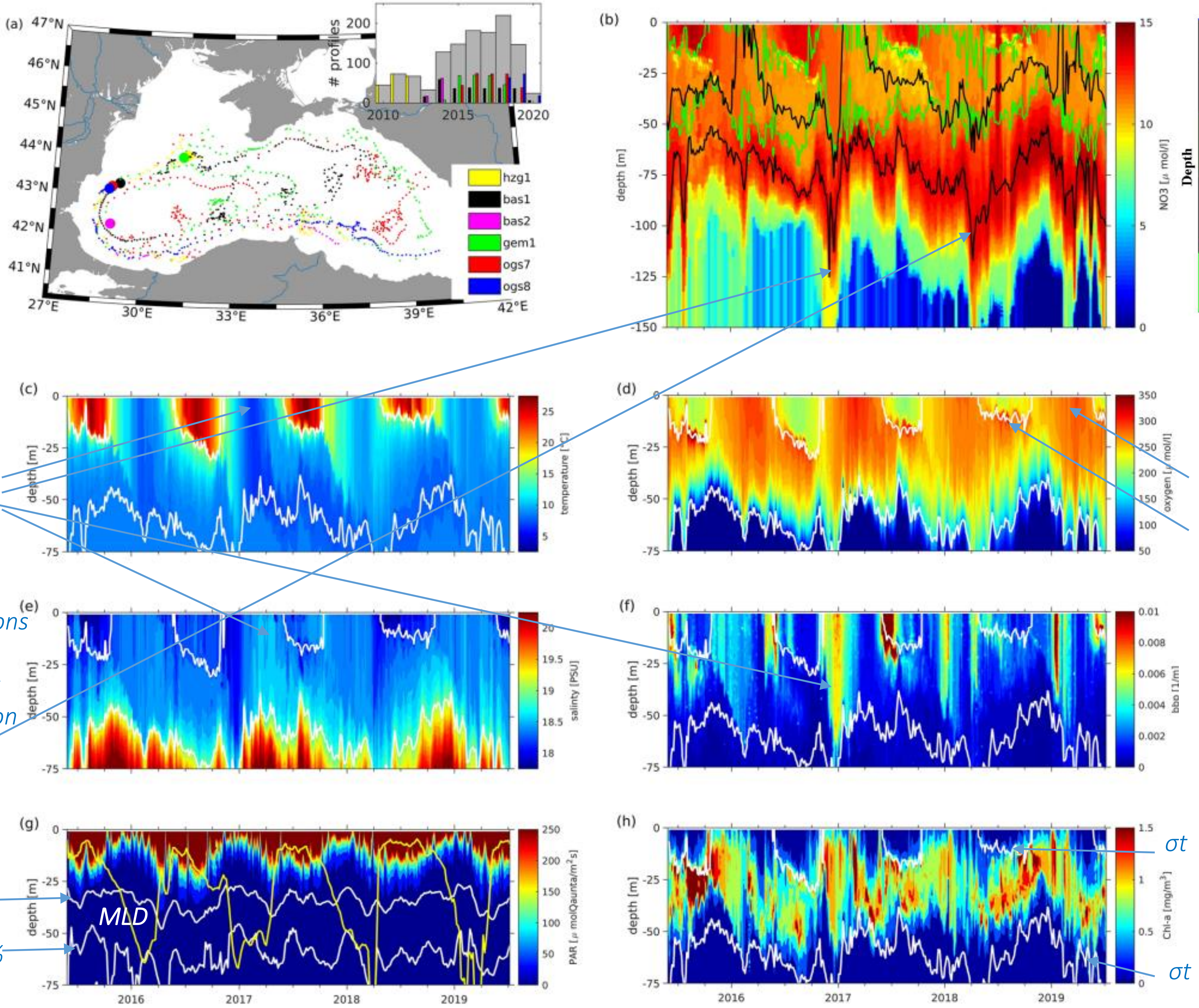


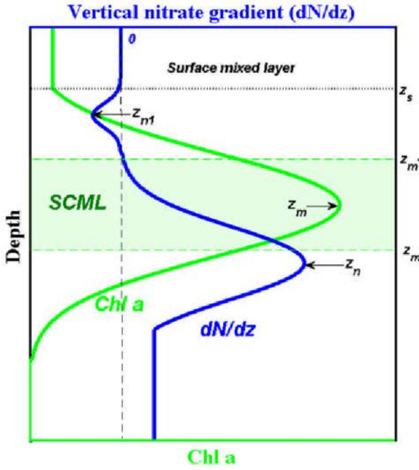


The key question is whether the currently available BGC-Argo data are sufficient to reliably describe the basin-wide spatial and temporal variability of BGC parameters, including their seasonal and interannual variations.

In Situ Data

Notice the overall low correlation among oxygen, bbp, and Chl-a





The year with the strongest cooling 2017

The range of density variations is explained as:
Above the base of the upper mixed layer, by T (illumination conditions)
Below: by S (dynamics of pycnocline)

1%

0.1%

Oxygen concentrations increase due to surface cooling and biological processes

particulate backscatter coefficient (bbp) *~ Coccolithophores, represented in the Black Sea mainly by Emiliania huxleyi*

σt = 12

subsurface chlorophyll maximum (SCM) layer

σt = 14.5.

Data From Numerical Models

Temperature, salinity, oxygen and Chl-a data produced by the Black Sea—Monitoring Forecasting Centre (BS MFC).

Physics: NEMO (version 3.6), implemented with horizontal grid resolution of 1/36° zonally and 1/27° meridionally (ca. 3 km) and 31 vertical levels.

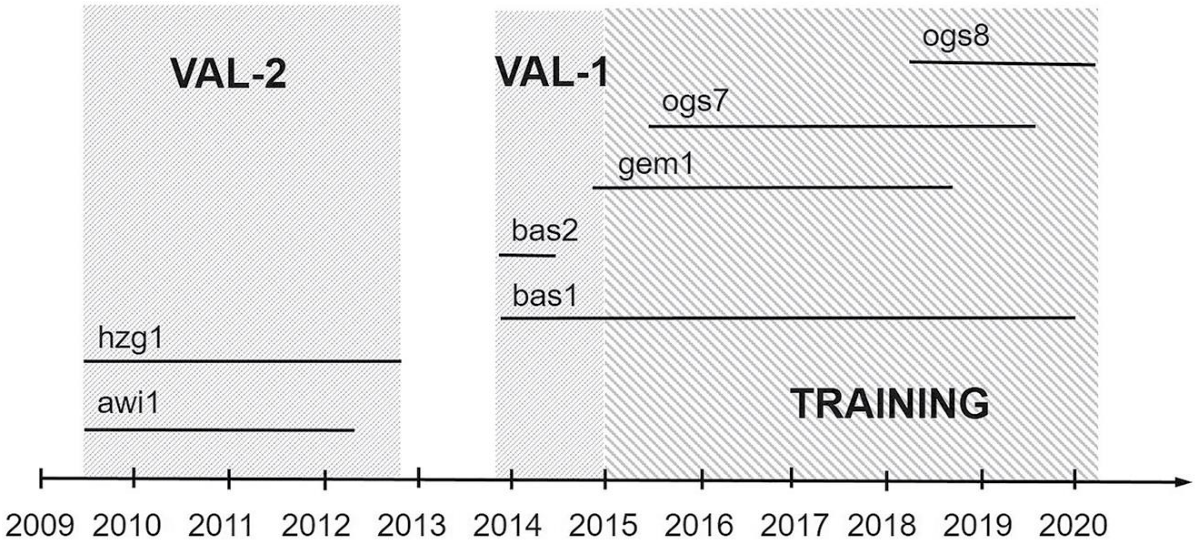
BGC: the BiogeochemicAl Model forhypoxic and Benthic Influenced areas (BAMHBI, Gregoire et al., 2008). The food web Is described through 24 state variables. BAMHBI is coupled with NEMO.

Feedforward Backpropagation Neural Network

Task: derive BGC variables from physical variables.

Essentially an input–output mapping in which the neurons combine the input data in such a way that the output can be considered a nonlinear combination of input data.

NN input: T and S
NN output: Chl-a, bbp and O_2



The training data set consists of 120,401 samples for Chl-a and bbp and 89,418 samples for oxygen. The validation data set is approximately 6 times smaller.

Index of agreement:

$$D(P,Q) = 1 - \sum_{i=1}^{n}(P_i - O_i)^2 / \sum_{i=1}^{n}\left(|P_i - \overline{O}| + |O_i - \overline{O}|\right)^2$$

Validation during the TRAINING period (bas1)

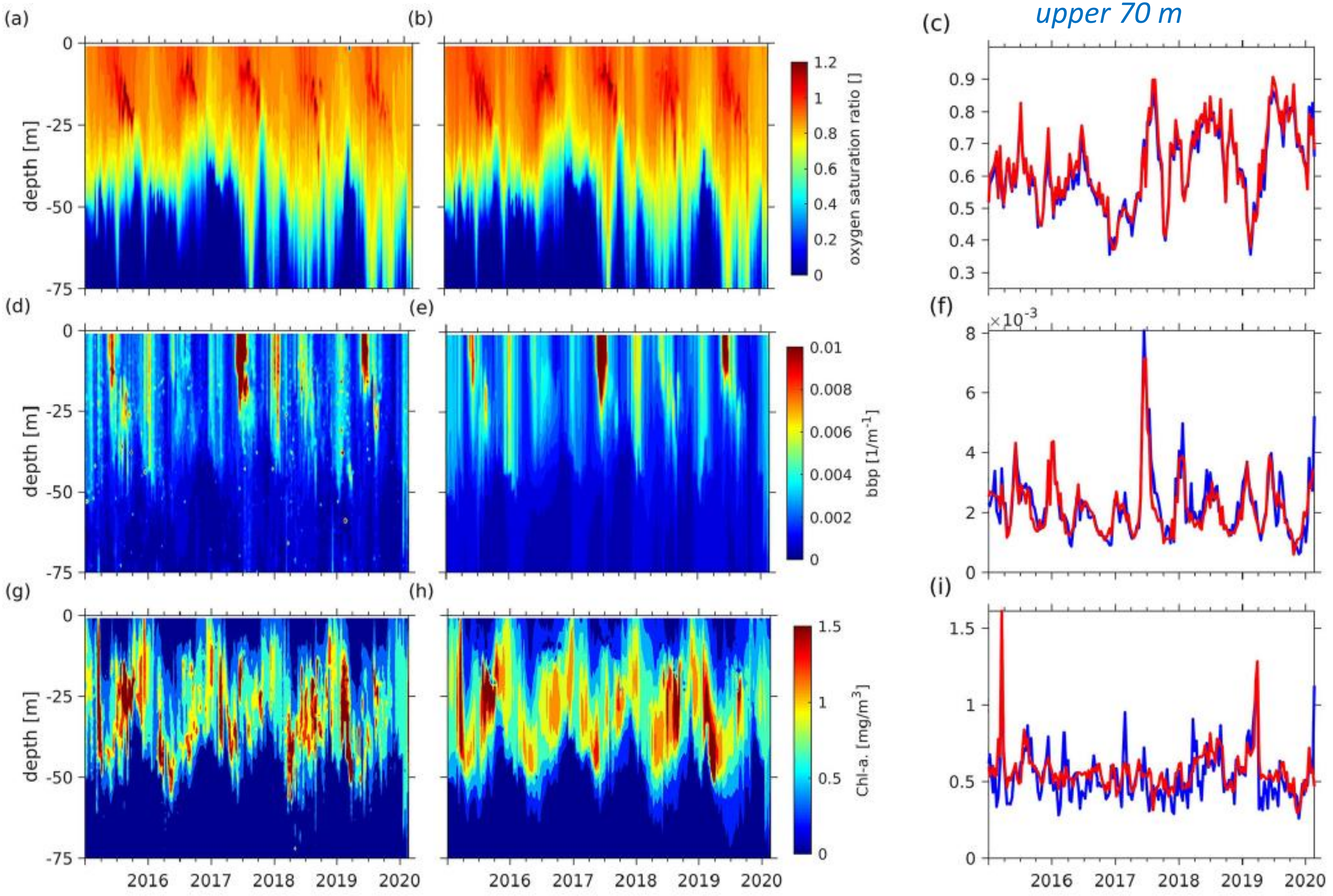
1 - perfect agreement
0 - no agreement at all
For the upper 70 m:

D=0.95

D=0.80

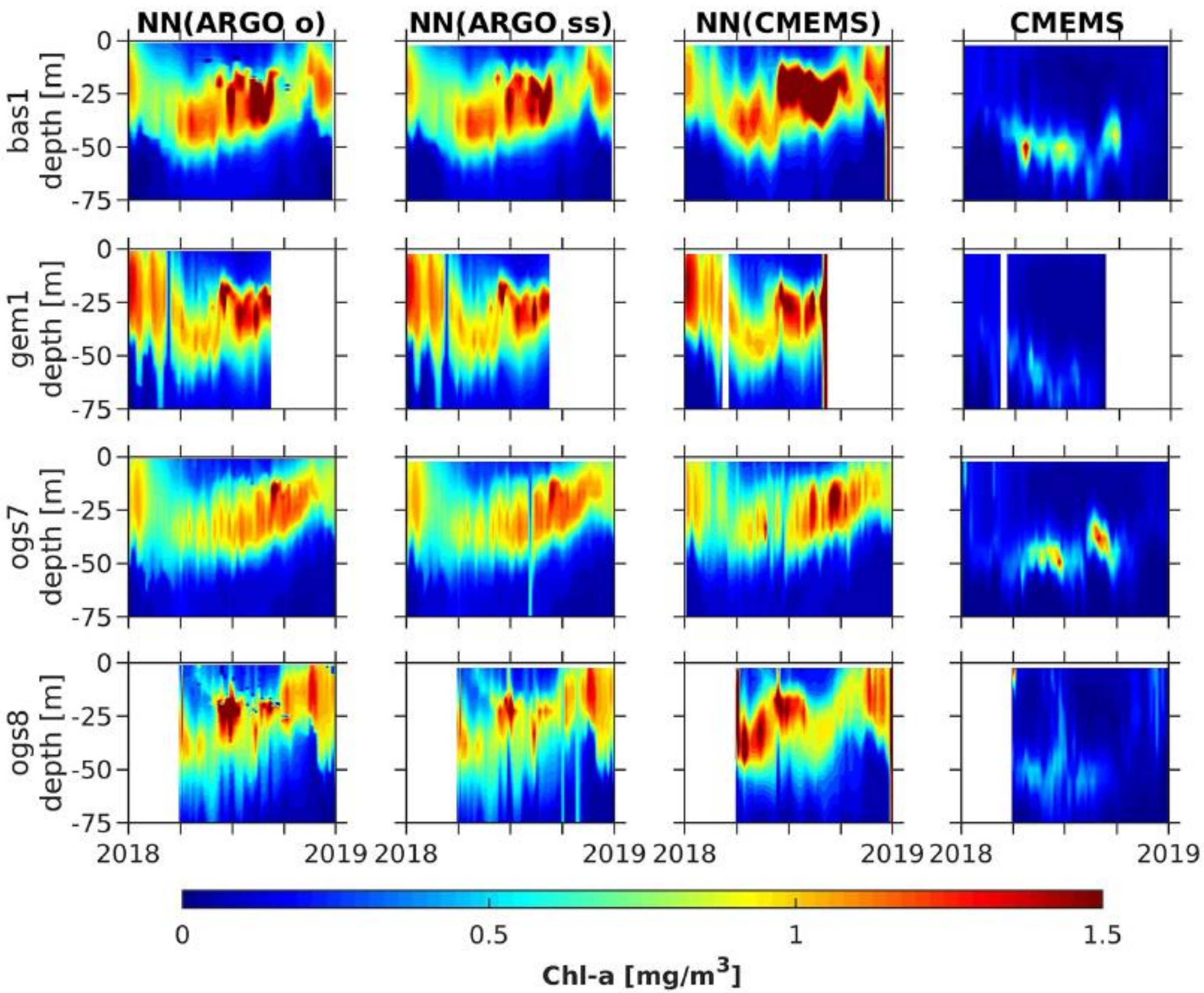
D=0.81

observed | emulated | Mean in the upper 70 m



Can we use the available data from numerical (physical) models as inputs for the developed NN to reconstruct the BGC states corresponding to the model physics?

A prerequisit of that is that the 3-D physical fields do not deviate much from the observations, This is true.



Mean Index of Agreement During 2018 Averaged Over the Upper 70 m Between the Observed BGC-Argo Float Data and the NN Reconstructions Derived Using Physical Data With a Different Origin or a Different Vertical Resolution as the Input

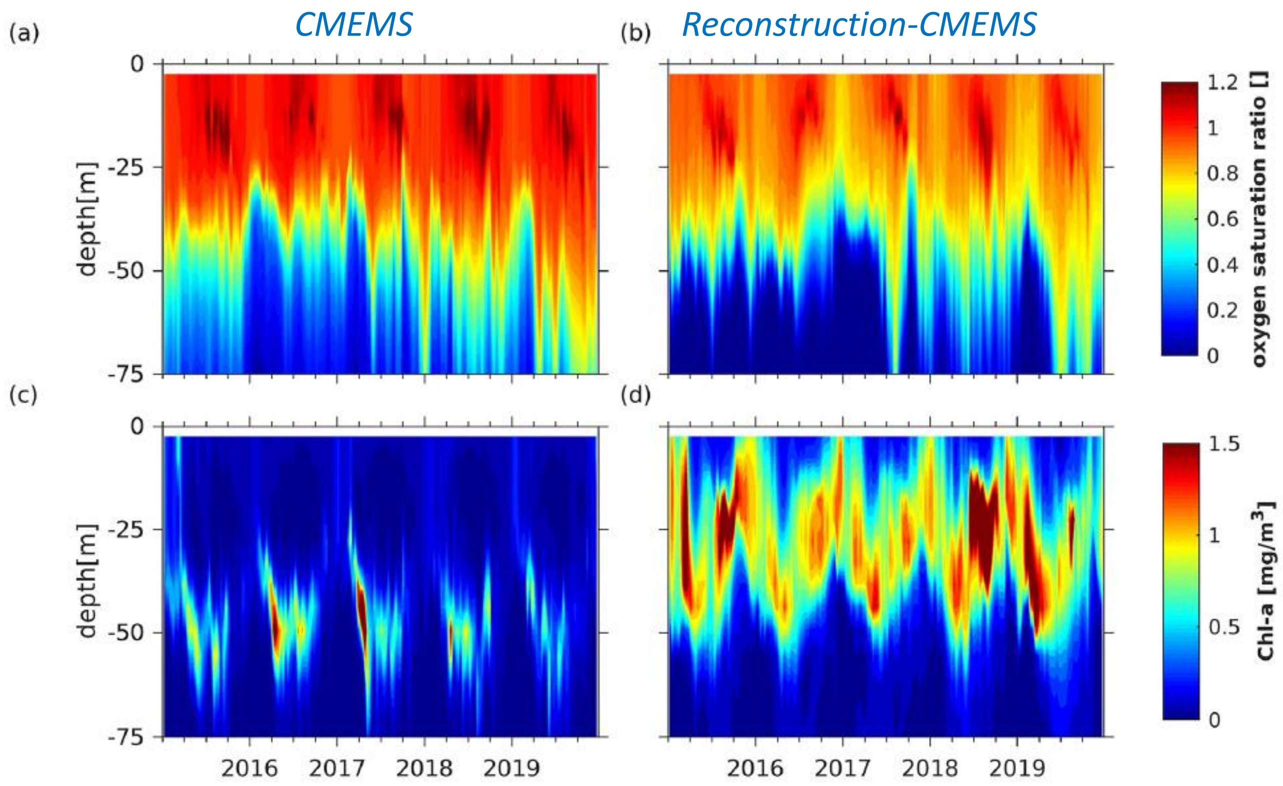
	NN(ARGO_o)	NN(ARGO_ss)	NN(CMEMS)	CMEMS (bio)
Oxygen				
bas1	0.92	0.92	0.64	0.42
gem1	0.92	0.92	0.66	0.51
ogs7	0.93	0.94	0.83	0.56
ogs8	0.90	0.90	0.62	0.40
bbp-				
bas1	0.73	0.73	0.54	-
gem1	0.92	0.92	0.81	-
ogs7	0.84	0.89	0.54	-
ogs8	0.77	0.77	0.51	-
Chl-a				
bas1	0.76	0.76	0.48	0.29
gem1	0.82	0.79	0.48	0.47
ogs7	0.71	0.79	0.53	0.25
ogs8	0.71	0.72	0.42	0.41

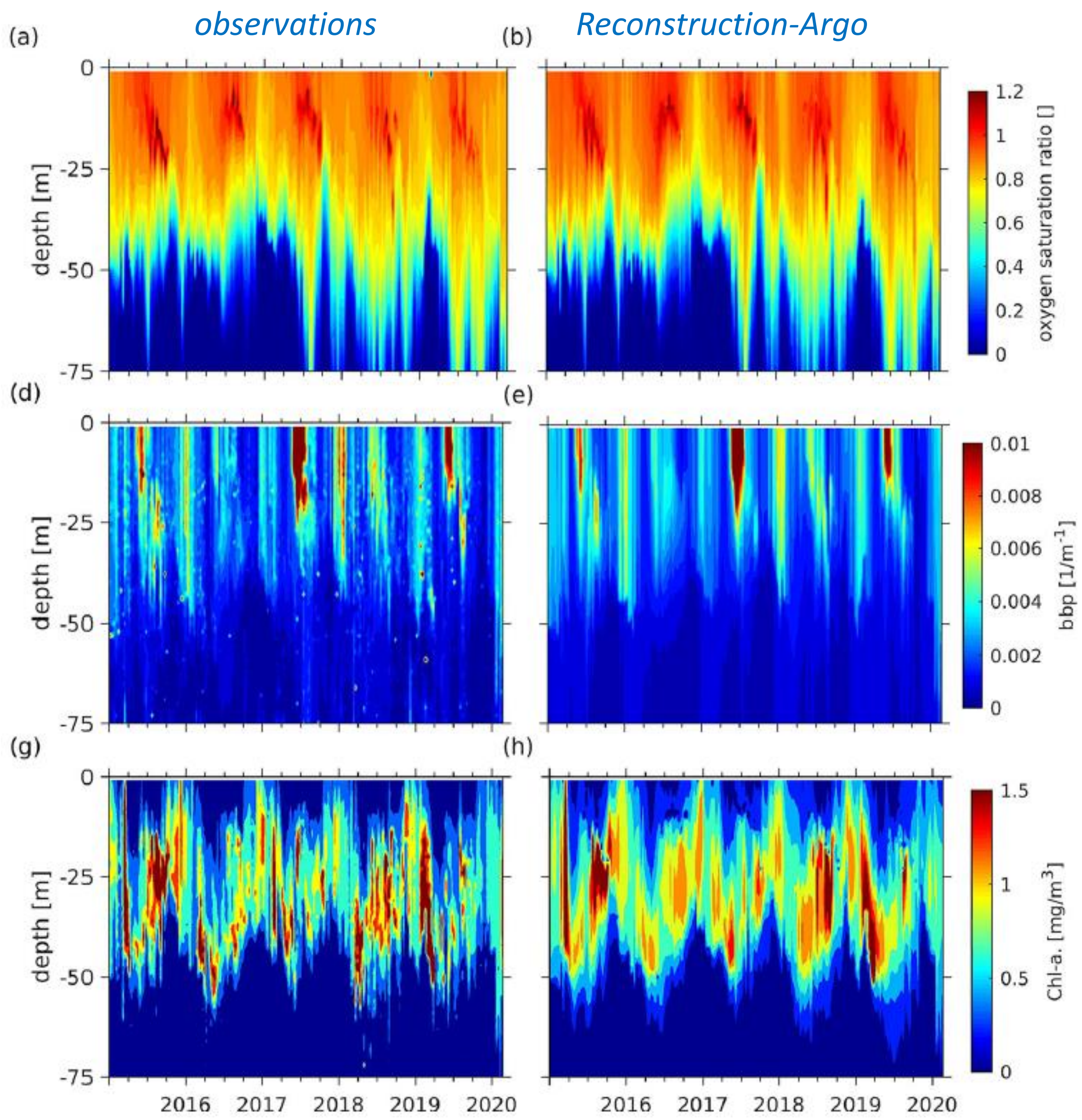
Note. The last column shows the index of agreement for the CMEMS BGC model data. Argo_o and Argo_ss denote the observed (o) and subsampled (ss) Argo data.

The NN reconstructions outperform the results of the model simulations

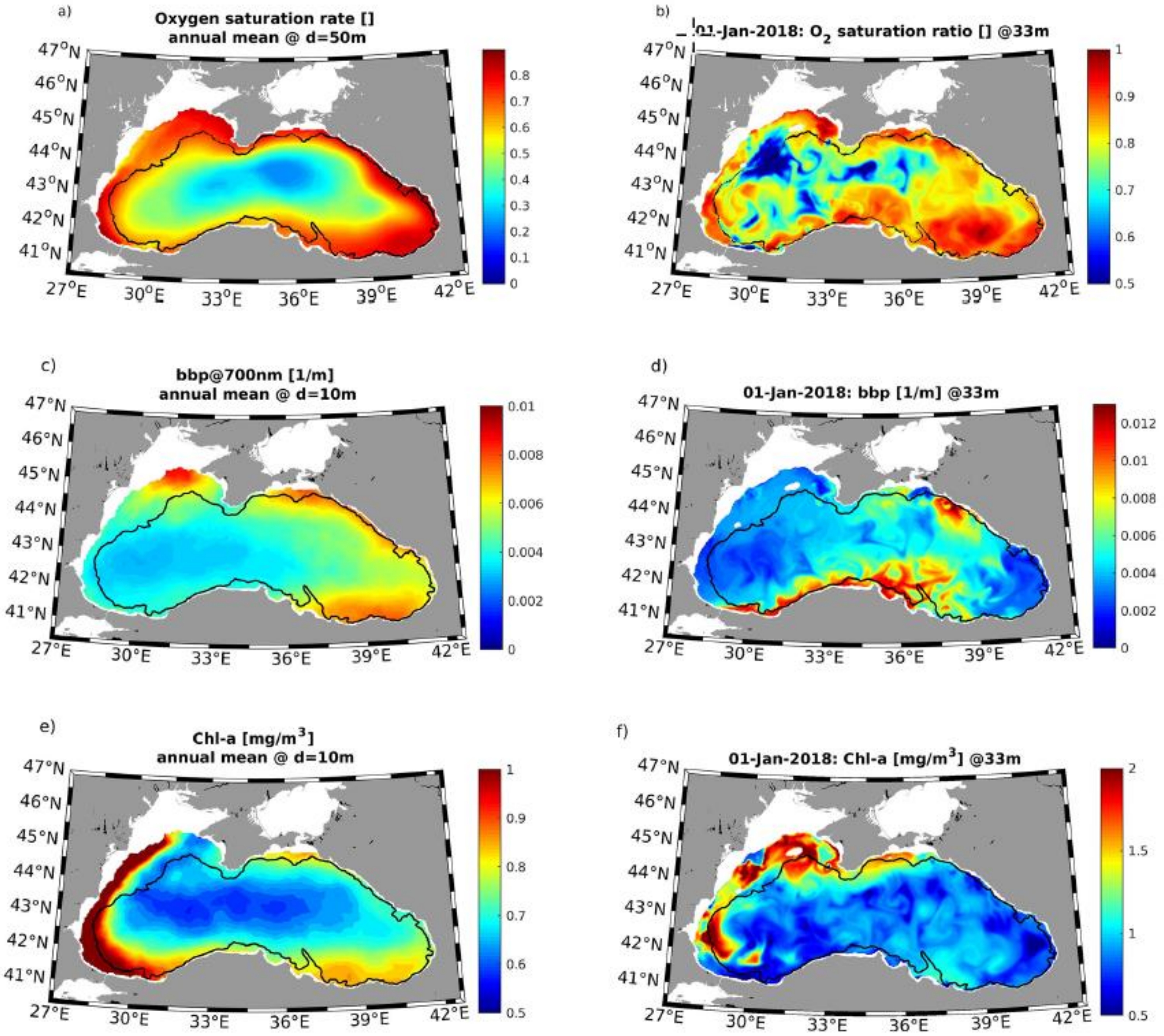
4D Reconstruction of BGC Dynamics-validation

float bas1

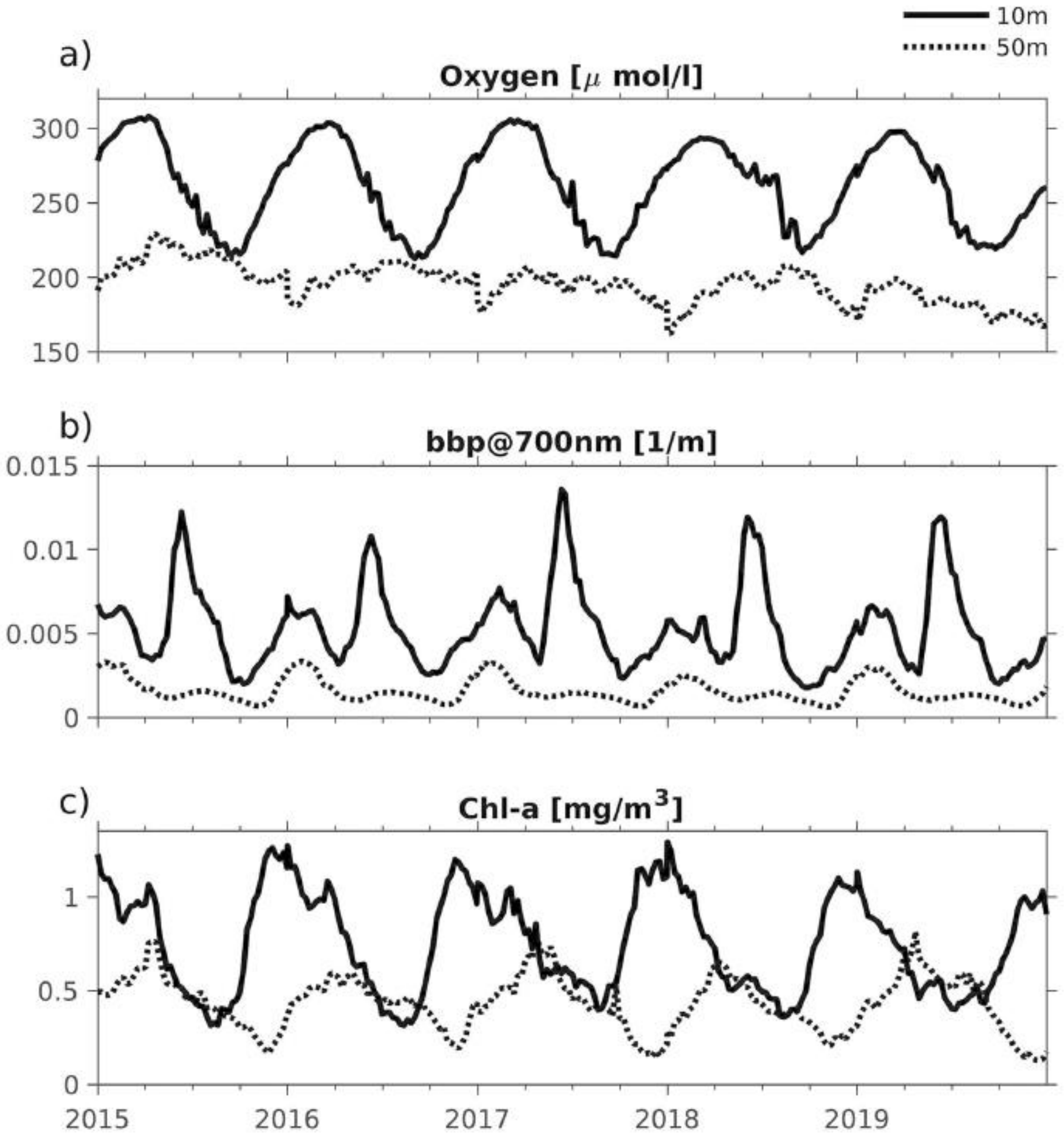




4D Reconstruction
Horizontal patterns



4D Reconstruction
Temporal variability



Conclusions

The magnitudes of responses of individual BGC parameters to physics appeared not to be proportional to the magnitude of physical variability.

ML helps to address the problem of undersampling by taking advantage of the synergy of different data sources and providing a tool to analyze the 4D structure of BGC fields.

There is perhaps a potential to enhance the quality of BGC numerical simulations by using ML (in data assimilation).

Physical data from numerical models appears to be a useful tool to replicate both the mean and eddy states of the Black Sea's BGC, as well as to identify the spatial patterns of recent BGC changes.

Stanev, E. V., Wahle, K., & Staneva, J. (2022). The synergy of data from profiling floats, machine learning and numerical modeling: Case of the Black Sea euphotic zone. Journal of Geophysical Research: Oceans, 127, e2021JC018012. https://doi.org/10.1029/2021JC018012