PML | Plymouth Marine Laboratory

Listen to the ocean

Deriving carbon pools on the North-West European Shelf through model-informed deep learning

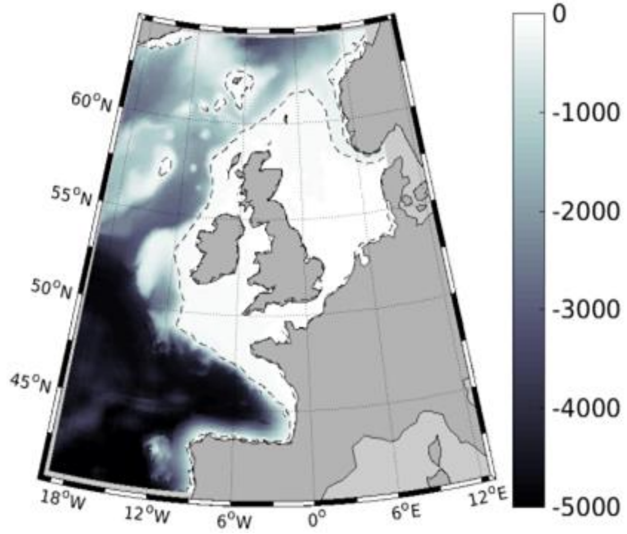
Jozef Skakala (PML, NCEO)

NECCTON
NEW COPERNICUS CAPABILITY FOR TROPHIC OCEAN NETWORKS

European Commission

UKRI Natural Environment Research Council

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL



Marine carbon cycle - NWES is an important region for carbon sequestration. Carbon cycles among different pools, e.g. phyto C, zoo C, DOC, detritus C, DIC, PIC and it is desirable to understand them, their variability and trends..

However..

- carbon-related NWES **in situ** data are quite sparse
- some (mostly global) **EO**-derived products are emerging, covering some C variables and fluxes (e.g. netPP, surface phyto carbon, ML-derived surface DOC…), but the comprehensive view is currently offered only by models

But… **models** have many significant biases and uncertainties, many of which are reduced through data-assimilation (DA). DA can become however very costly and difficult, if we want to update model space far from the observed variables. E.g. our NWES system constrains phyto C near the ocean surface, but not much more…

Main idea: *assume from the model equations/dynamics emerges a map/function between observable variables and carbon stocks. Assume* ***(big if!)*** *that this map is correct, i.e. that model C biases would originate more from the inputs to the map than from the map itself..*

We could then learn the map through ML using model simulations and then use it to expand the observational products to carbon pools!

Advantages:

- a lot of comprehensive training data
- model specifically adapted to NWES
- we can predict any variable covered by the model

Disadvantages:

- the big assumption!

Observable variables: SST, surface salinity, PFT chlorophyll, structural (lat, lon, day, bathymetry), atmospheric (SWR, WS) and riverine (river deposition of bgc tracers and runoff)
- all taken at the same time than predicted variables

Predicted variables:

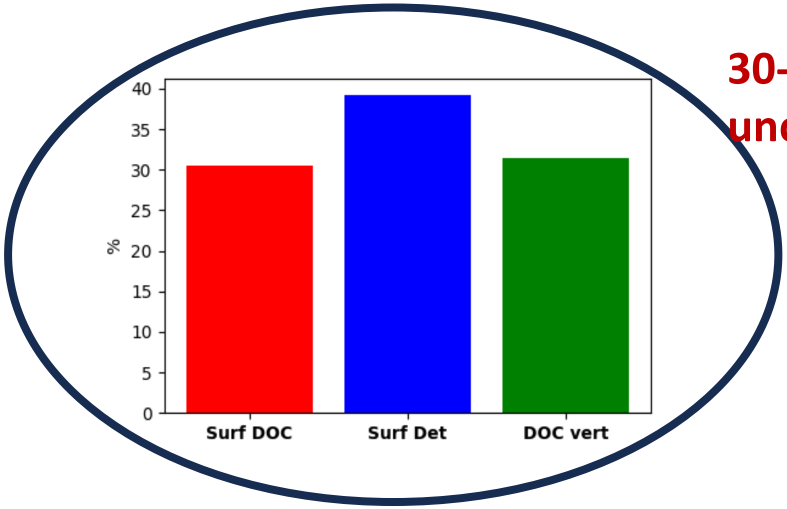
- Surface total DOC, detritus, zooplankton, bacteria carbon and DIC
- Vertically averaged DOC, bacteria C and detritus C beneath 50m

We focus on NEMO-FABM-ERSEM NWES operational model.. Is the ERSEM obs -> C map "correct"? All we can say is how the model seems to be representing detritus C and DOC.. The limited validation indicates that simulated detritus is fine, whilst DOC has very large (Summer) biases.. There is current ongoing NECCTON work on improving those biases..

A question:

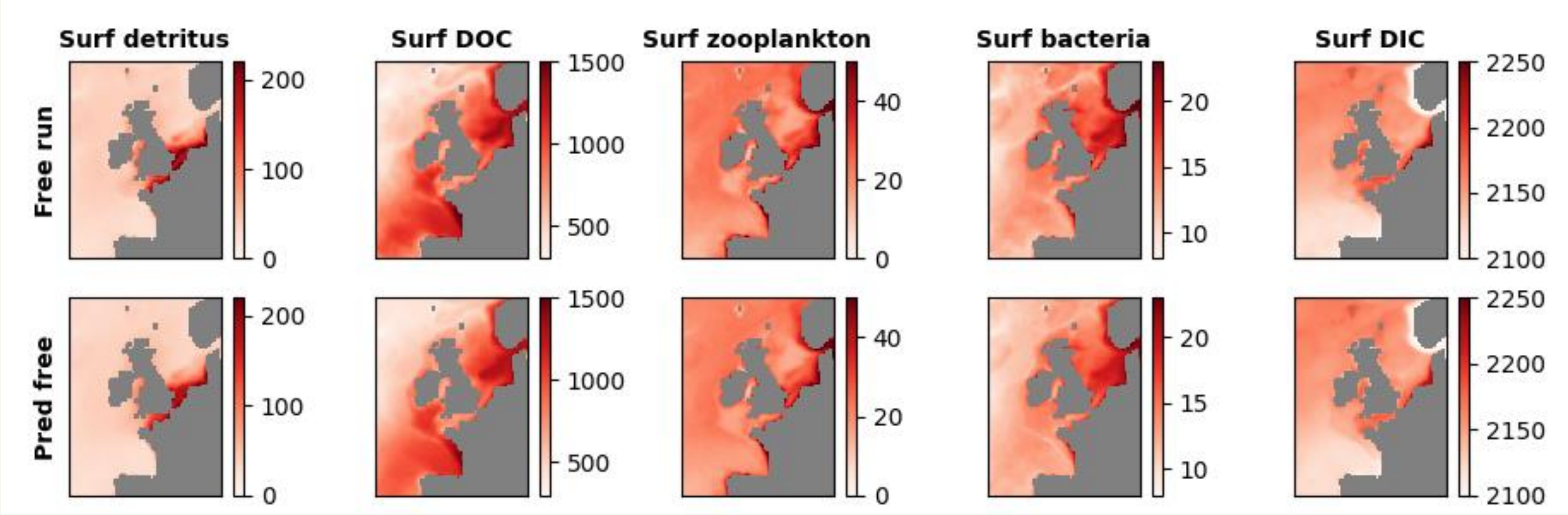
Model and therefore the obs → C map are uncertain.. **Can we estimate how uncertain this map is?**

Done through a 1D 90-member ensemble at L4 – model uncertainty is represented by the variability in dominant parameters.



30-40% uncertainty

How well does the NN model perform with validation data?

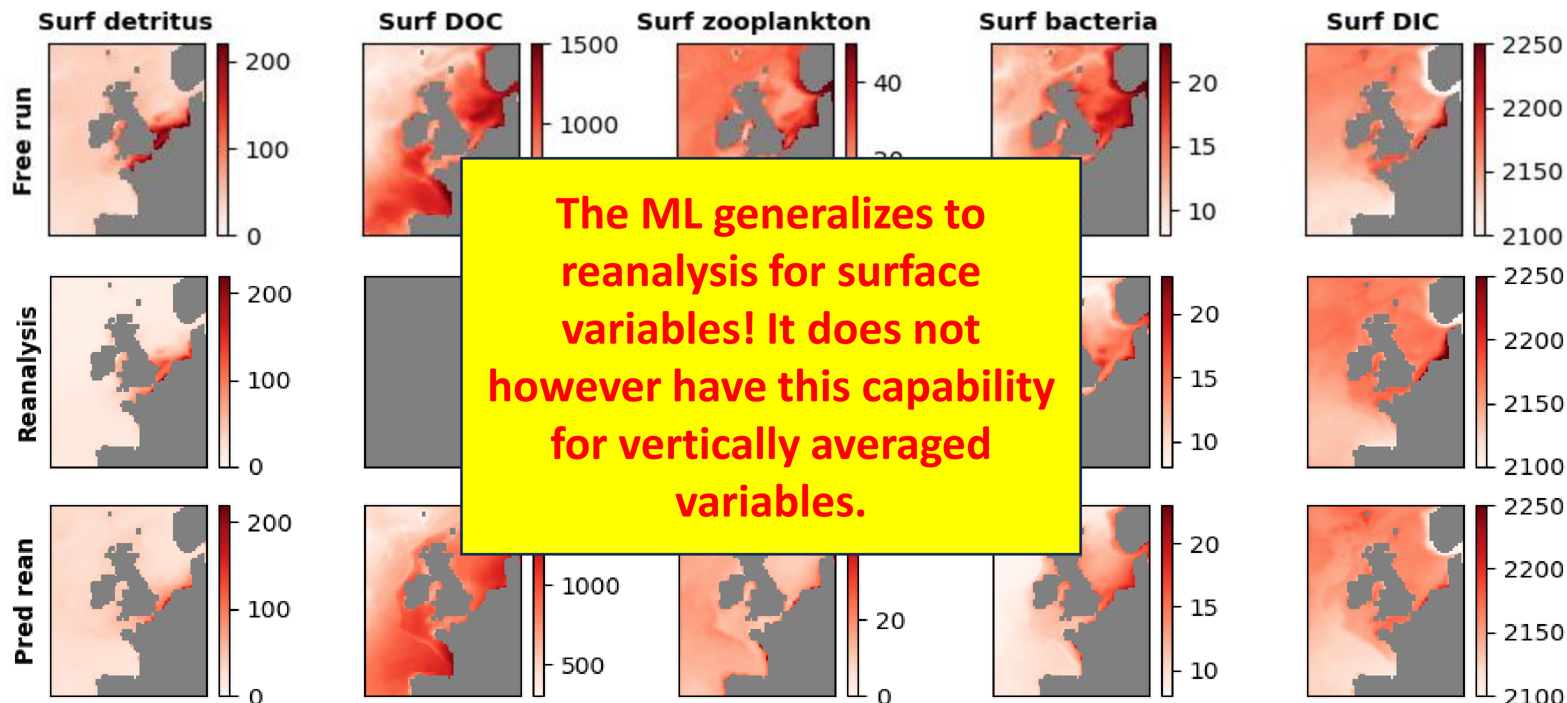


The skill to predict test data is very good!

Surface	Detritus C	DOC	Zooplankton C	Bacteria C	DIC
R^2	0.82	0.83	0.85	0.8	0.88

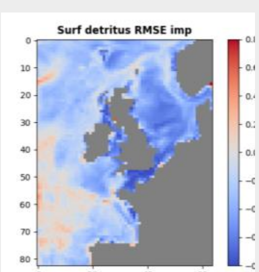
Vertical	DOC	Bacteria C	DIC
R^2	0.9	0.72	0.91

How does the model generalize, i.e. what happens if we supply instead of model inputs EO data? *I.e. changing SST and PFT chlorophyll inputs..*

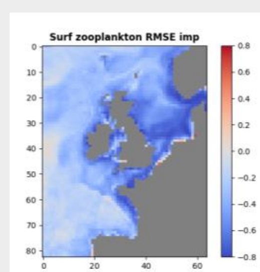


The ML generalizes to reanalysis for surface variables! It does not however have this capability for vertically averaged variables.

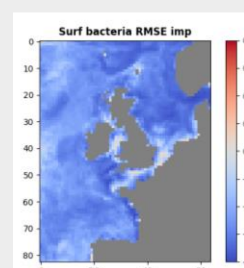
Detritus



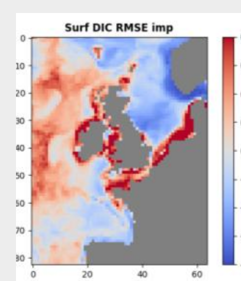
Zooplankton



Bacteria

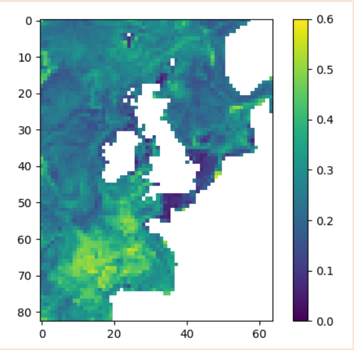


DIC

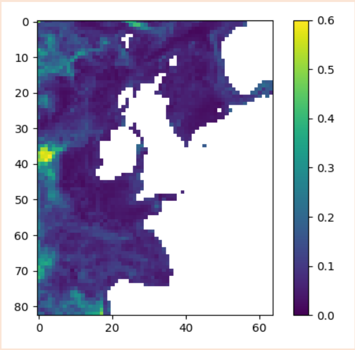


Vertically averaged variables do not generalize, i.e. they have long (climatological) time-variability and can be sufficiently predicted from structural variables.

Detritus surf

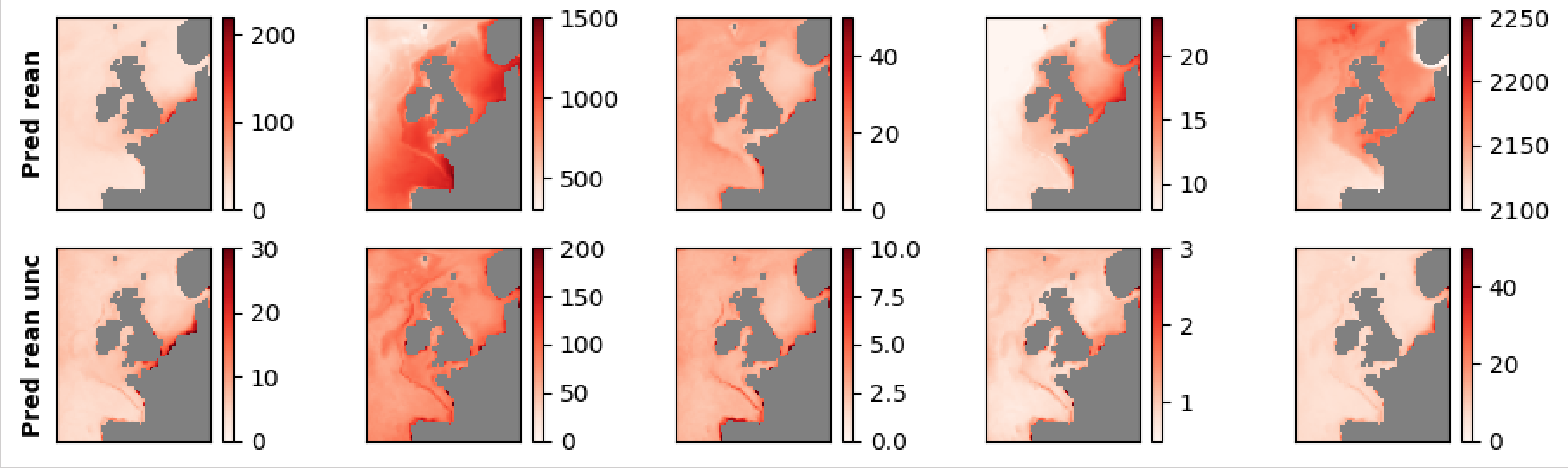


Detritus vert



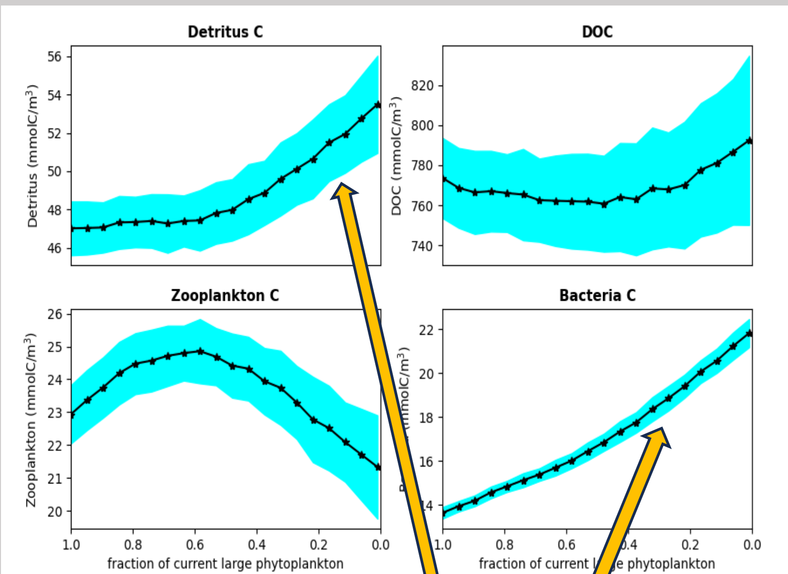
Uncertainty quantification

Detritus | DOC | Zooplankton | Bacteria | DIC



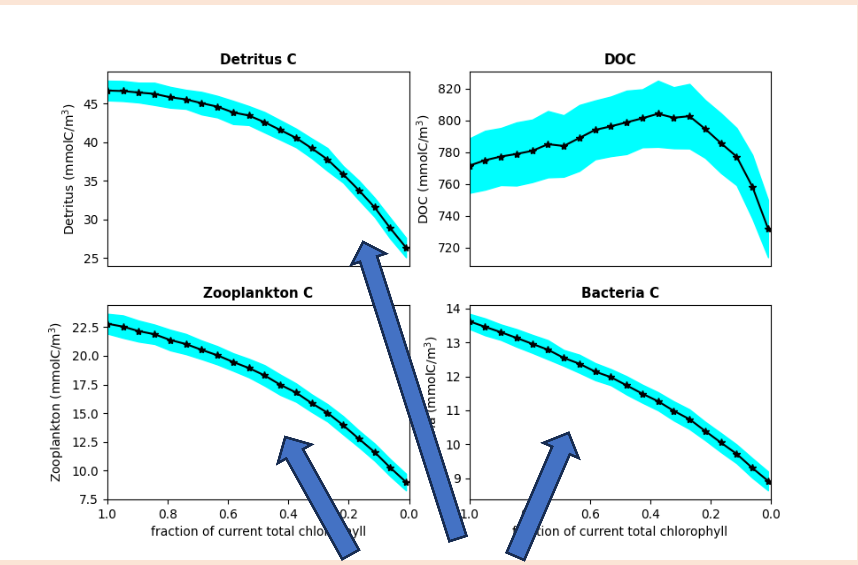
Can we explore uncertainty-assessed what-if scenarios?

How does change in phytoplankton community structure impact carbon export?



Less sinking of detritus from the surface with the decrease of species size (less carbon export). Increase of decomposers

How does biogeochemistry change with reduced primary production?



Reduction of higher trophic levels and detritus with reduced primary production.

Conclusions

- deep learning can be developed to learn the model map between observable variables and carbon pools. The map can be successfully used with observations as inputs, therefore emulating the dynamical adjustment that happens in the reanalysis.

- we have shown some feasibility of running what-if "climate" scenarios using the NN models. Although they were uncertainty-assessed, this is clearly limited.. Their biases should be better understood, advanced and better ways of testing their realism need to be applied.

- after additional testing these tools can be used to efficiently produce long term (e.g. past 1998) carbon products deriving them from atmospheric/physics reanalyses and/or observations.

- these tools can be also deployed within DA (e.g. through balancing scheme) to improve reanalyses and operational forecasts by making them more multi-variate

Thank you!