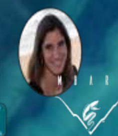
L'IA POUR RECONSTRUIRE DES SÉRIES LONGUES ET COMPRENDRE LA VARIABILITÉ BASSE FRÉQUENCE DE LA BIOMASSE PHYTOPLANCTONIQUE DANS L'OCÉAN GLOBAL

ET COMPRÉHENSION DES MÉCANISMES SOUS JACENTS



E. Martinez, T. Gorgues
E. Pauthenet, M. Roux

M. Lengaigne

R. Fablet, L. Dumetz
M. Lakra

O. Pannekoucke

M. Sourisseau

D. Raitsos

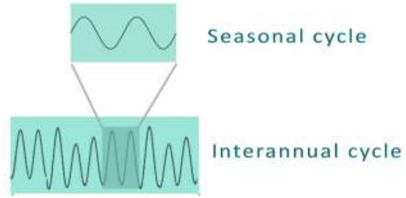
M. Messié

Atelier commun LEFE-CYBER / ILICO / ODATIS. (4 & 5 juin 2024, PARIS, LOCEAN)

Utilisation de l'IA pour analyse de données issues de séries longues

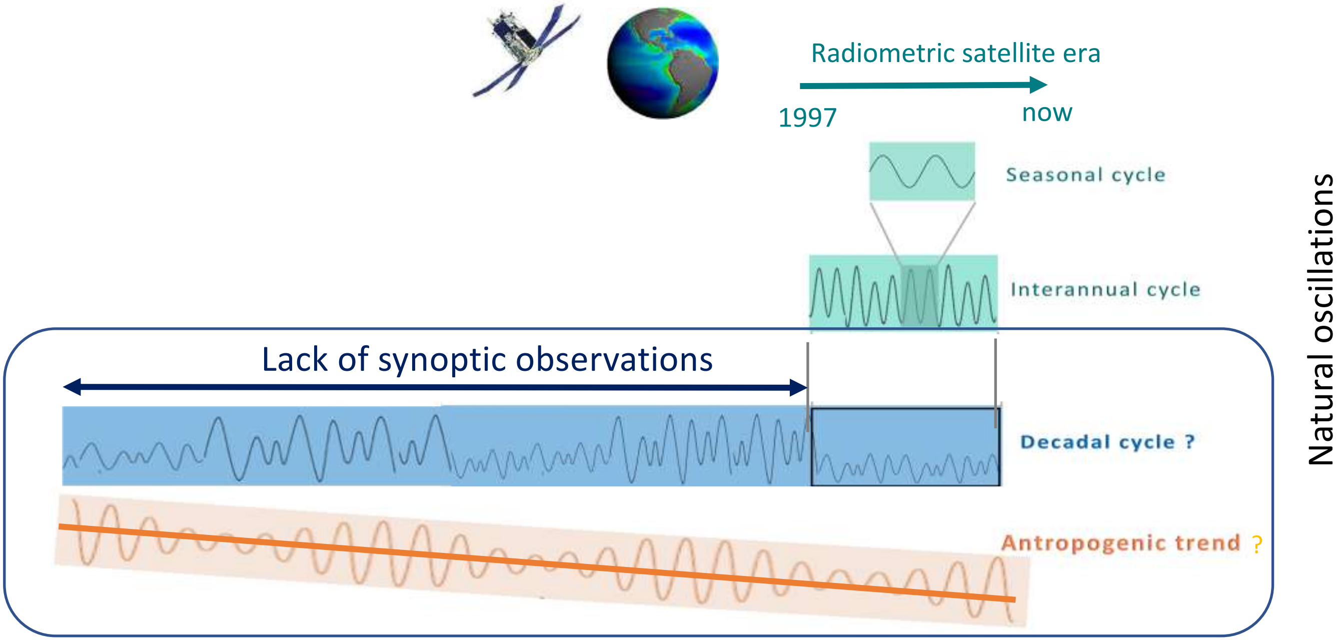
Radiometric satellite era

1997 now



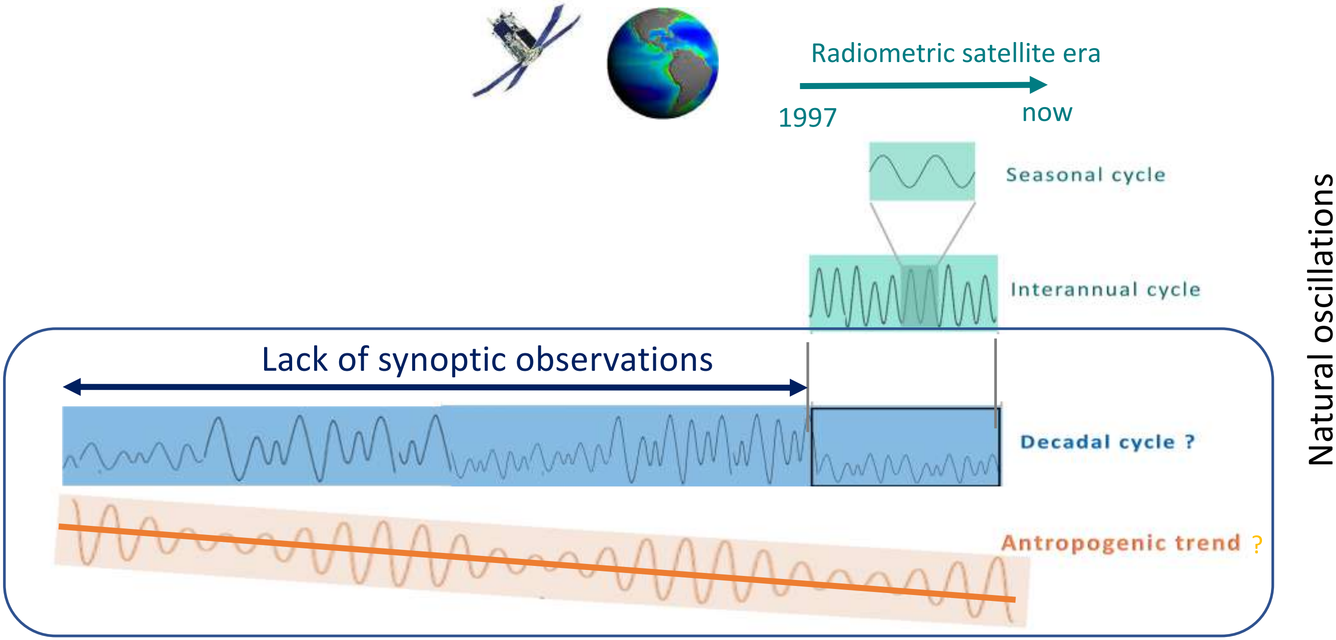
Natural oscillations

?



scientific lock 1:

- Too short satellite obs. time-series

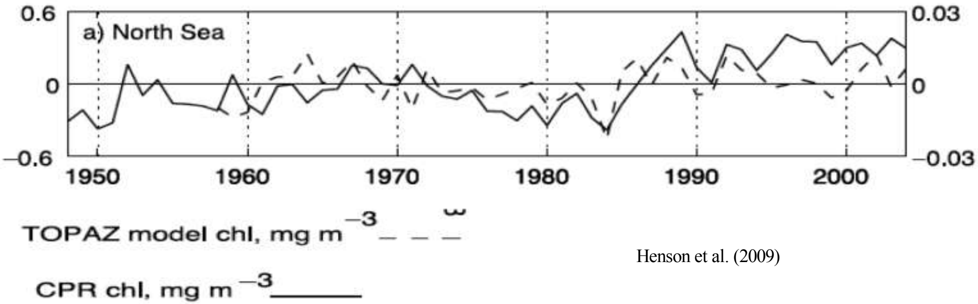


scientific lock 1:

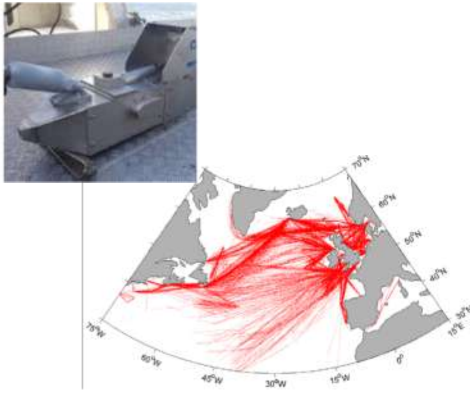
- Too short satellite obs. time-series

Aim 1:

- Identify the natural oscillations vs. anthropic trend
- Understand the underlying BGC-physical processes



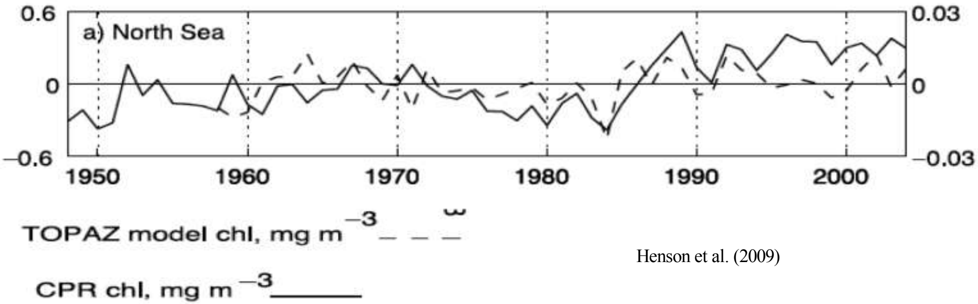
Henson et al. (2009)



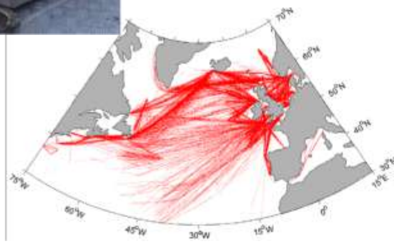
Continuous Plankton Recorder (CPR) since 1960

scientific lock 2:

- Difficulty in parameterizing biology in BGC models (i.e., multi-decadal and regime shifts)



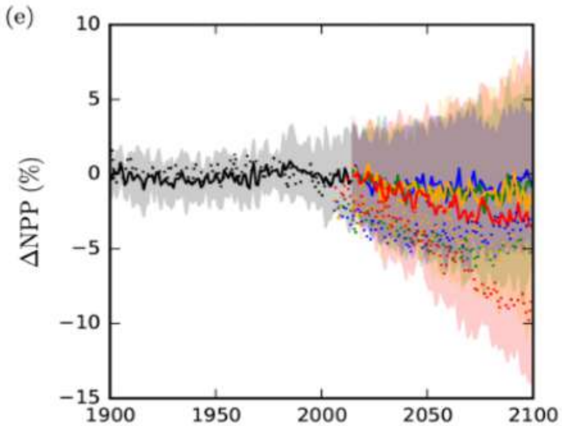
Henson et al. (2009)



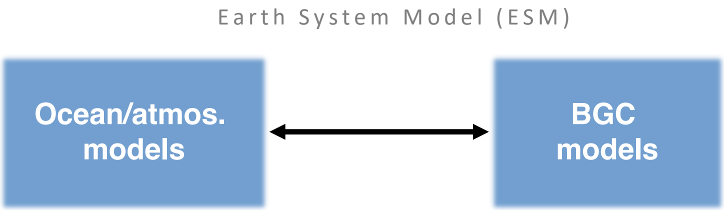
Continuous Plankton Recorder (CPR) since 1960

scientific lock 2:
- Difficulty in parameterizing biology in BGC models (i.e., multi-decadal and regime shifts)

Aim 2:
- Better understand biotic and abiotic interactions on the phyto- & zoo plankton biomass and communities

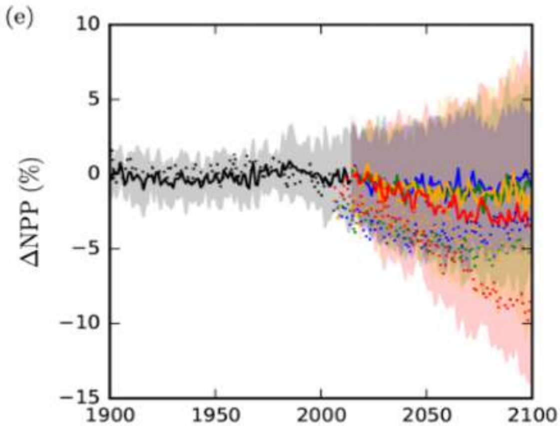


Global mean projections of depth-integrated net primary production (%)
(from Kwiatkowski et al., 2020)

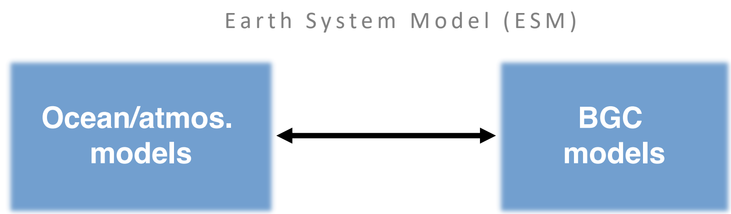


scientific lock 3:

- Wide cone of uncertainty on climate projections



Global mean projections of depth-integrated net primary production (%)
(from Kwiatkowski et al., 2020)



scientific lock 3:

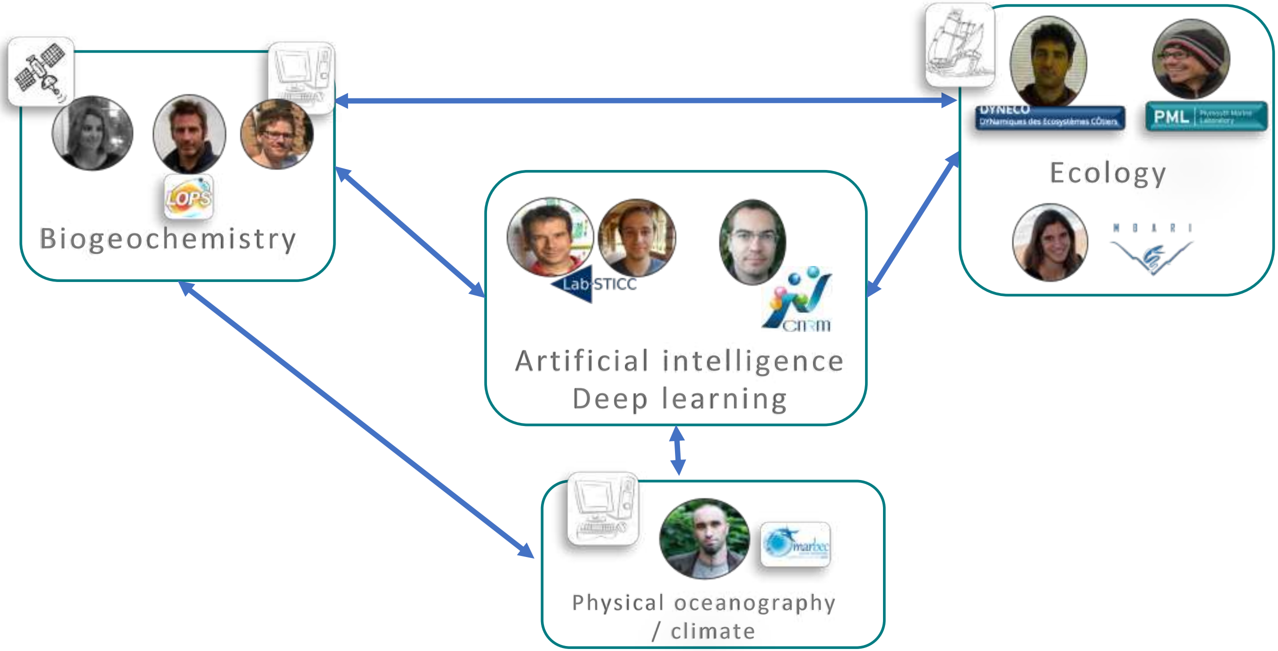
- Wide cone of uncertainty on climate projections

Aim 3:

- Quantification of ESM's uncertainty related to physical forcing & to biogeochemical model formulation
- Identification of physical processes

Deep leaRning approaches to Elucidate phytoplAnktonic cliMate induced variability (ANR PRC DREAM, 2023-27)

anr agence nationale de la recherche
AU SERVICE DE LA SCIENCE

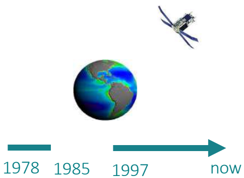


agence nationale de la recherche
AU SERVICE DE LA SCIENCE

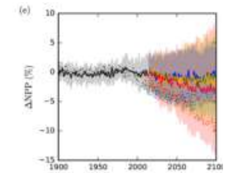
Aim:

- assess multi-decadal variability & trends of phytoplankton biomass
- Understand the underlying physical and BGC processes

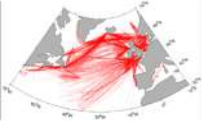
WP2a: reconstruct past multi-decadal reconstruction at global scale



WP3: quantify physical & bgc uncertainties in climate models



WP2b: better understand the link between phyto and zooplankton biomass and communities

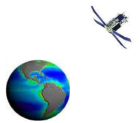


Deep leaRning approaches to Elucidate phytoplAnktonic cliMate induced variability (ANR PRC DREAM, 2023-27)

Aim:

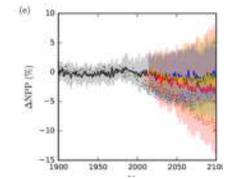
- assess multi-decadal variability & trends of phytoplankton biomass
- Understand the underlying physical and BGC processes

WP2a: reconstruct past multi-decadal reconstruction at global scale



1978 1985 1997 now

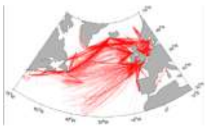
WP3: quantify physical & bgc uncertainties in climate models



WP1 Improve the deep learning architectures through various emulators

- Data-driven: Y= f(X)
- Ordinary and Partial Differential Equations: $\partial_t Y = f(\partial_t X)$, $\partial_t Y = f(\partial_t X, \partial_t Y)$....

WP2b: better understand the link between phyto and zooplankton biomass and communities

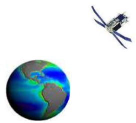


Deep leaRning approaches to Elucidate phytoplAnktonic cliMate induced variability (ANR PRC DREAM, 2023-27)

Aim:

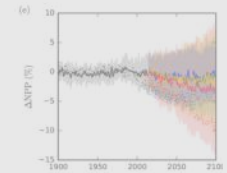
- assess multi-decadal variability & trends of phytoplankton biomass
- Understand the underlying physical and BGC processes

WP2a: reconstruct past multi-decadal reconstruction at global scale



1978 1985 1997 now

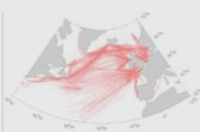
WP3: quantify physical & bgc uncertainties in climate models

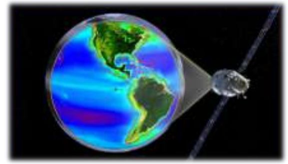


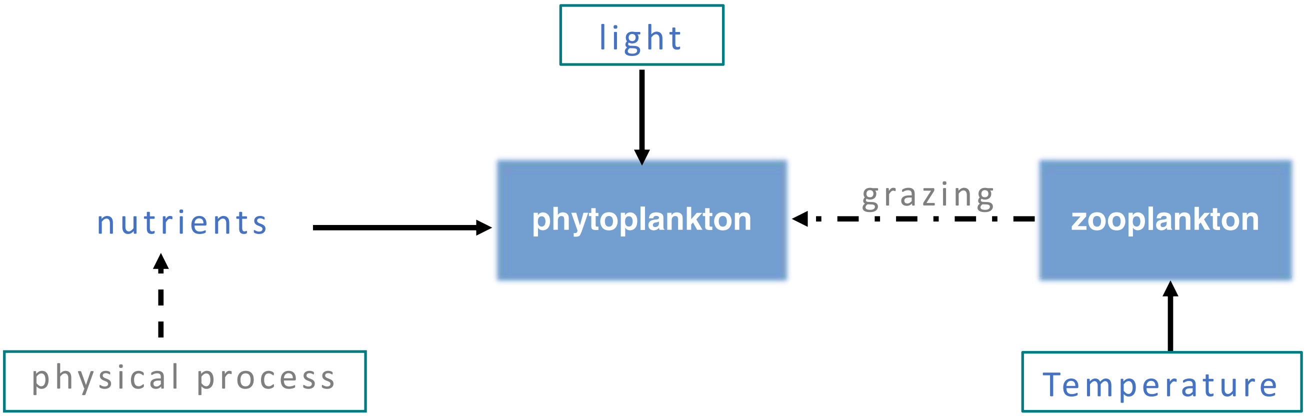
WP1 Improve the deep learning architectures through various emulators

- Data-driven: Y= f(X)
- Ordinary and Partial Differential Equations: $\partial_t Y = f(\partial_t X)$, $\partial_t Y = f(\partial_t X, \partial_t Y)$....

WP2b: better understand the link between phyto and zooplankton biomass and communities





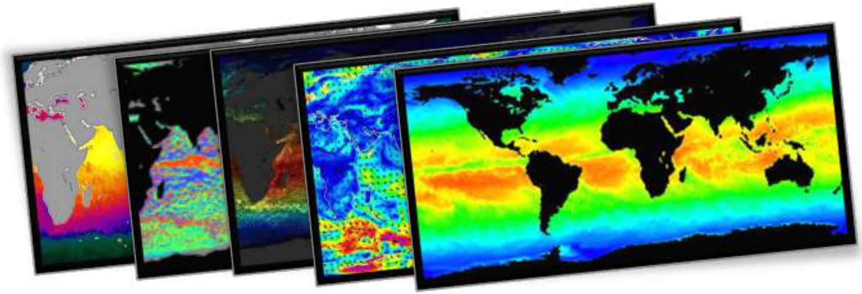


Hypothesis: physically driven at global scale

ML/DL & radiometric observations to reconstruct long time series

Sea surface temperature
Sea level anomaly
Surface currents (zon. & merid.)
Surface winds (zon. & merid.)
Short wave radiations

geophysical predictors:

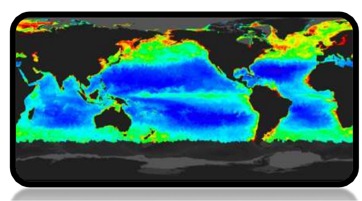


1998

1 TRAIN

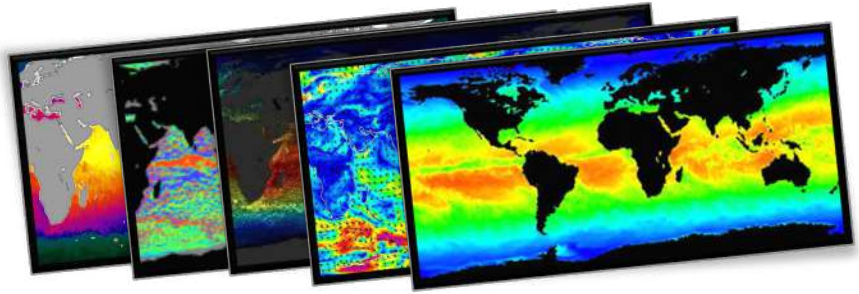
Numerical Schemes

1 target:
Chl



Sea surface temperature
Sea level anomaly
Surface currents (zon. & merid.)
Surface winds (zon. & merid.)
Short wave radiations

geophysical predictors:



1900's

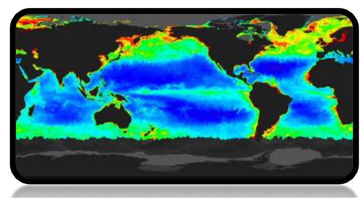
1998

1 TRAIN

2 APPLY

Numerical Schemes

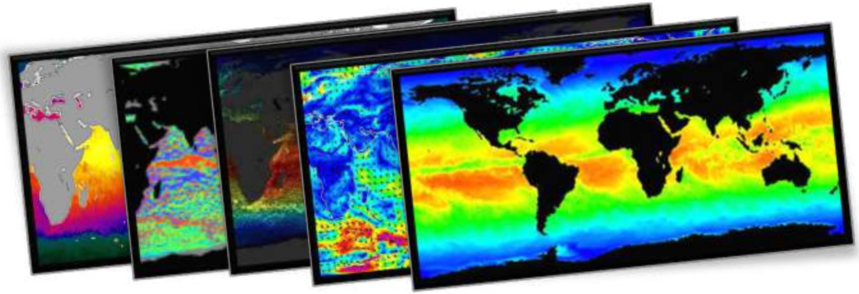
1 target:
Chl



ML/DL & radiometric observations to reconstruct long time series

Sea surface temperature
Sea level anomaly
Surface currents (zon. & merid.)
Surface winds (zon. & merid.)
Short wave radiations

geophysical predictors:



1900's

1998

1 TRAIN

2 APPLY

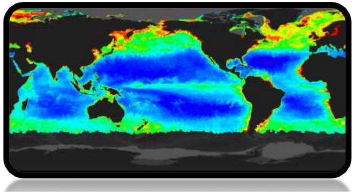
2b

RECONSTRUCT

Numerical Schemes

→ $\mathrm{Chl}_{\mathrm{Reconstructed}}$
Data analysis

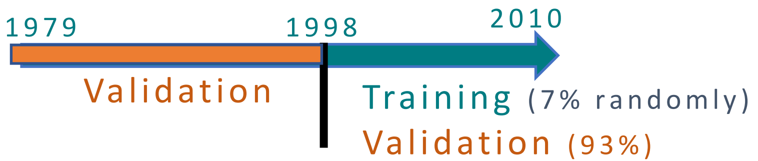
1 target:
Chl



A) The proof of concept on a physical-BGC model (SVR)

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010

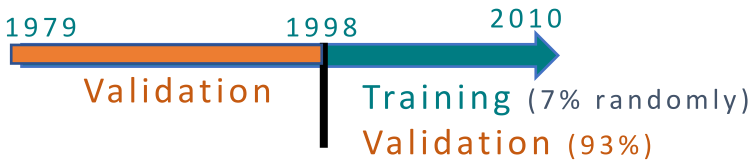
1°x1°grid monthly
Predictors:
7 + lon, lat, time



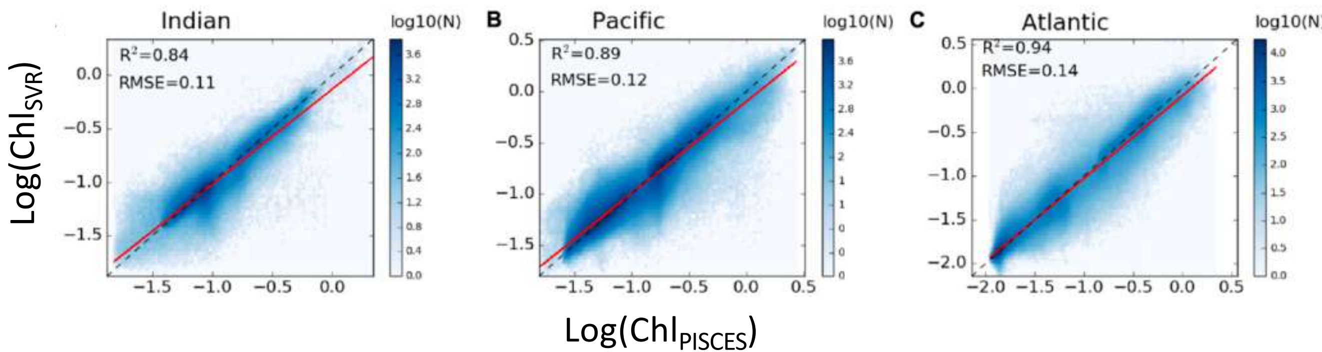
A) The proof of concept on a physical-BGC model (SVR)

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time



1979-1997



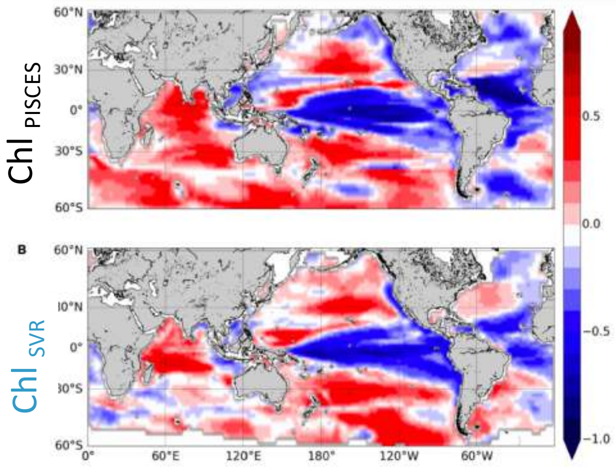
Martinez et al., Frontiers in Marine Sciences 2020

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Ability to reproduce the low frequency varibility

1st mode of interannual EOF:



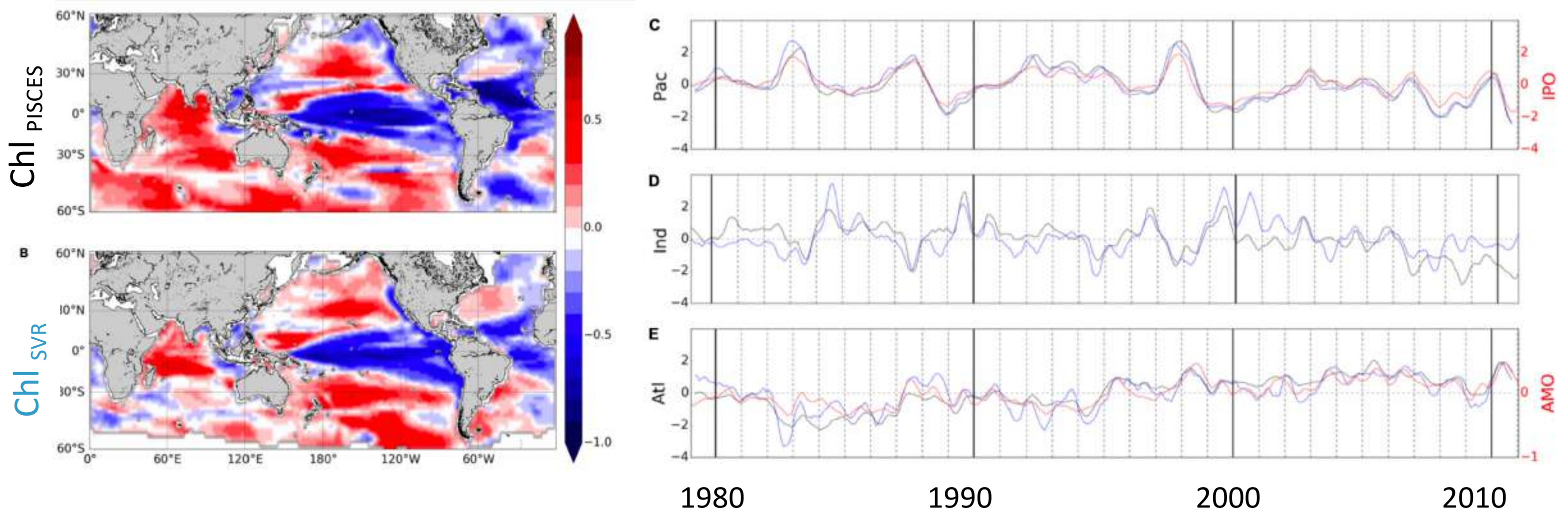
Martinez et al., Frontiers in Marine Sciences 2020

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Ability to reproduce the low frequency varibility

1st mode of interannual EOF:

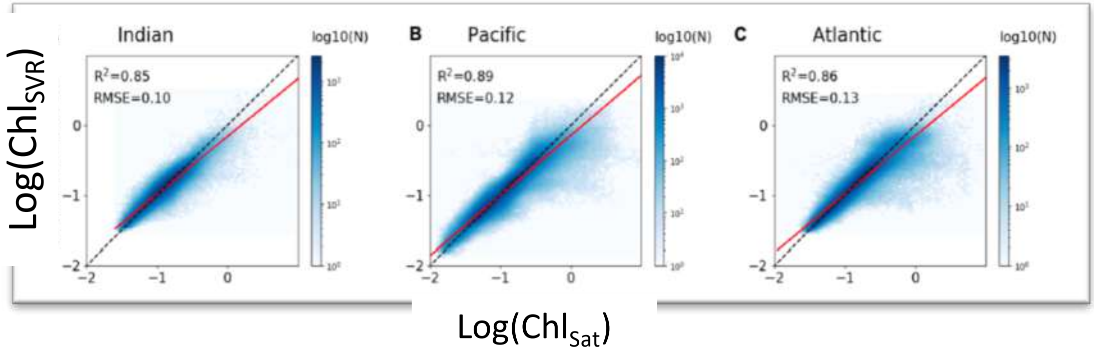


Martinez et al., Frontiers in Marine Sciences 2020

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010
SVR	NEMO-PISCES model	Satellite (OC-CCI)	1998-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Training: 7% randomly
Validation: 93%



- Capture seasonal, interannual variability and trends
- **Regional biais**

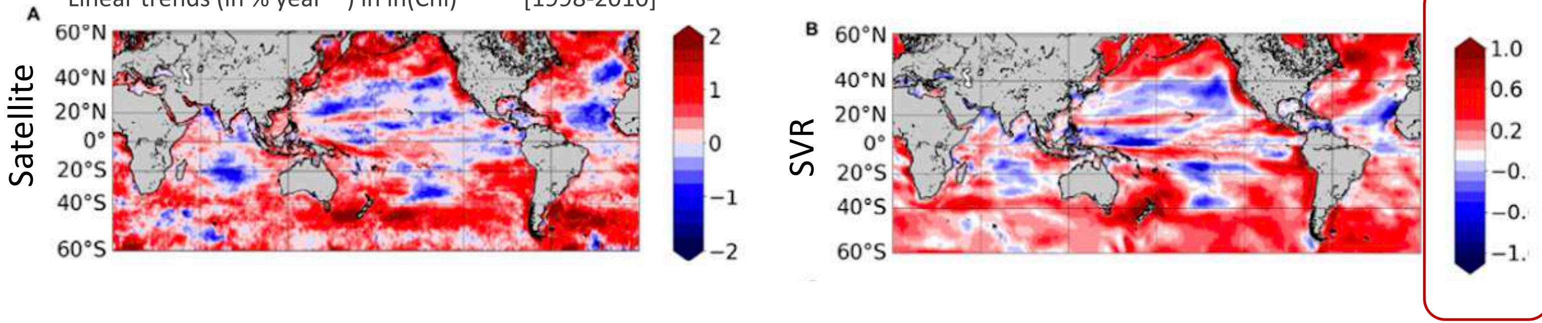
Martinez et al., Frontiers in Marine Sciences 2020

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010
SVR	NEMO-PISCES model	Satellite (OC-CCI)	1998-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Training: 7% randomly
Validation: 93%

Linear trends (in % year $^{-1}$) in ln(Chl) [1998-2010]



- Capture seasonal, interannual variability and trends
- **Regional biais**
- **Underestimation of reconstructed Chl**

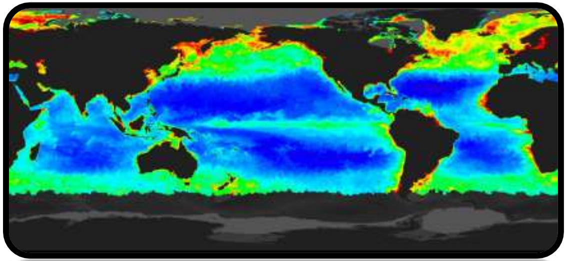
Martinez et al., Frontiers in Marine Sciences 2020

Aside: we don't want to use longitude, latitude as predictors

→ boundaries of BGCP are not static

Reygondeau et al., journal of BGC 2013

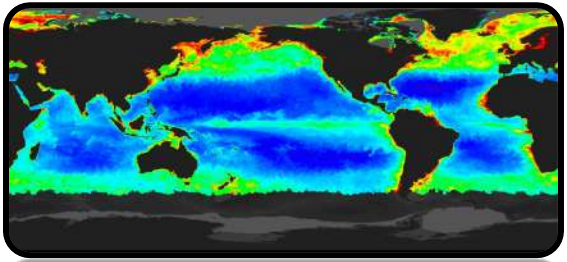
Chl



Aside: we don’t want to use longitude, latitude as predictors
→ boundaries of BGCP are not static

Reygondeau et al., journal of BGC 2013

Chl



Multi-Layer Perceptron (MLP) vs. SVR:

- MLP slightly more skillfull than SVR (Chl underestimation & spatial biais remain)
- MLP reconstruction slighly improved when trained with 80% of the data
- MLP is less, but still sensitive to longitude, latitude, as predictors

→ To overcome these limitations …….

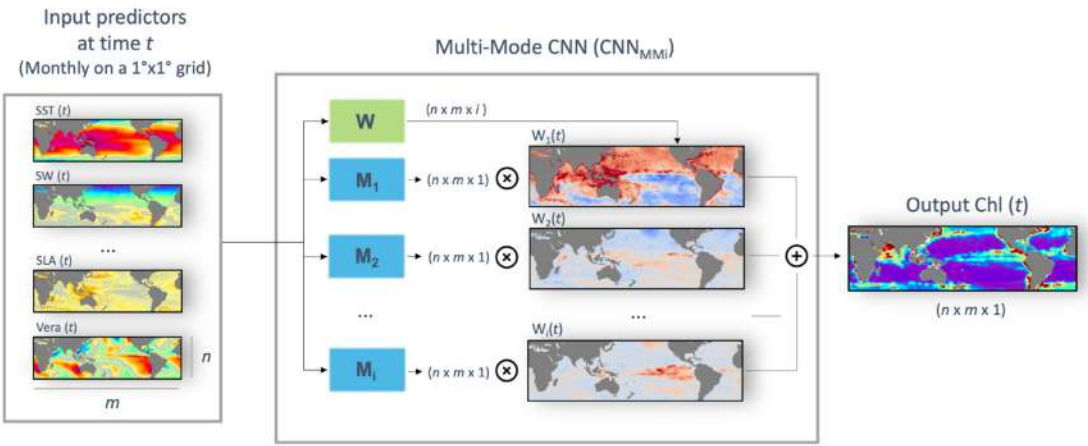
key features of the multi-mode CNN architecture:

- can explicitly account for regional/spatial physics-driven variabilities
- no need to a priori delineate BGCPs boundaries
- space-time activation of each mode to improve the interpretability of the network.

Roussillon et al., Frontiers in Marine Sciences 2023

key features of the multi-mode CNN architecture:

- can explicitly account for regional/spatial physics-driven variabilities
- no need to a priori delineate BGCPs boundaries
- space-time activation of each mode to improve the interpretability of the network.



CNN with 8 modes: optimum of performance and interpretability

Roussillon et al., Frontiers in Marine Sciences 2023

D) Getting insights into physical processes (multi-mode CNN)

Predictors: from satellite obs & reanalysis

Target: satellite OC

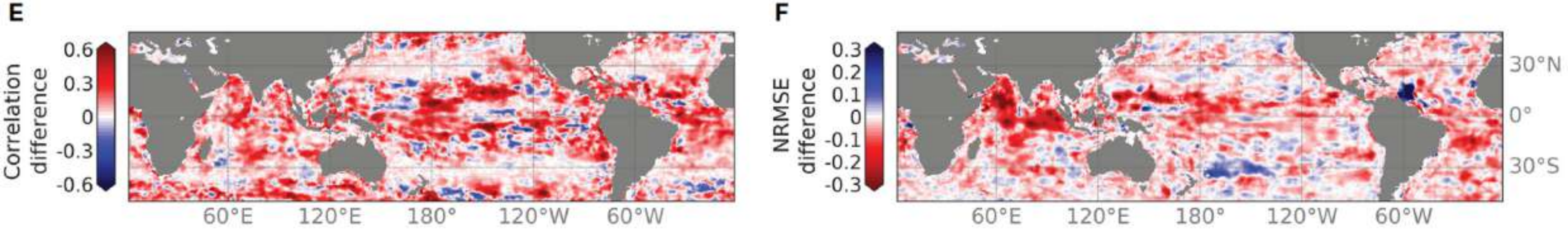
1°x1°grid monthly

Roussillon et al., Frontiers in Marine Sciences 2023

Test 2012-2015:

Predictors		Global scatterplot		
	Model	R^2	RMSE	Slope
9 (+ bathymetry + continental binary mask)	CNN_1	0.84	0.31	0.85
	$\mathbf{CNN_{MM8}}$	**0.87**	**0.28**	**0.90**

Difference CNN_{MM8} vs. CNN_1 when compared to satellite Chl

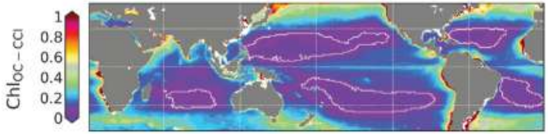


Modes of CNN_{MM8} can regionally learn specific phytoplankton responses to the physical forcing

→ better capture some regional processes than CNN_1

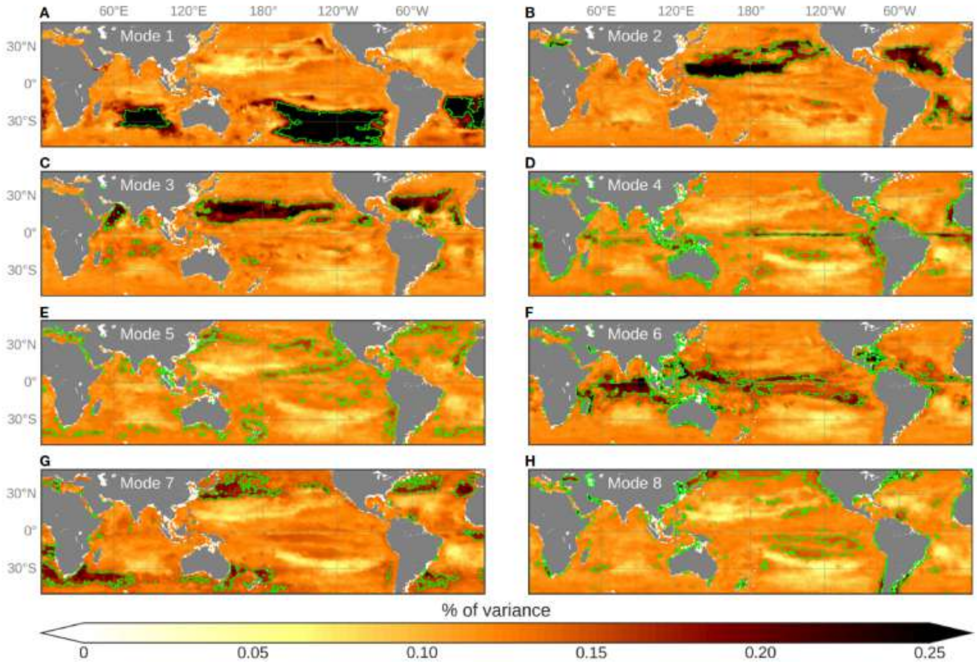
Roussillon et al., Frontiers in Marine Sciences 2023

D) Getting insights into physical processes (multi-mode CNN)



Consistent and relevant regions appear

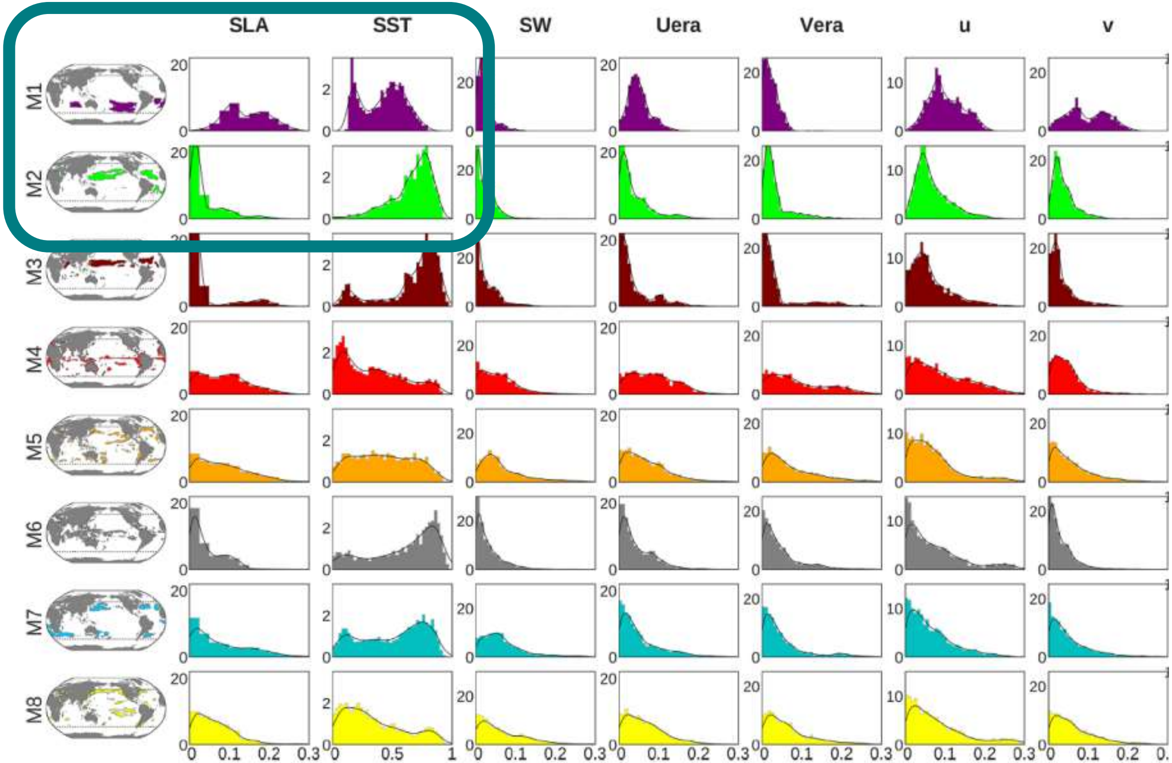
Percentage of variance explained by each 8 modes of CNN_{MM8}. Isolines of percentile-90 of the values are superposed in green.



Roussillon et al., Frontiers in Marine Sciences 2023

D) Getting insights into physical processes (multi-mode CNN)

Normalized distribution (y-axis) of the relative importance of the 7 physical predictors (x-axis) computed over the percentile-90 area for each mode.



NB: the scale on the x-axis is homogeneous across all the variables but SST.

Roussillon et al., Frontiers in Marine Sciences 2023

Key features from the « usual architecture » :

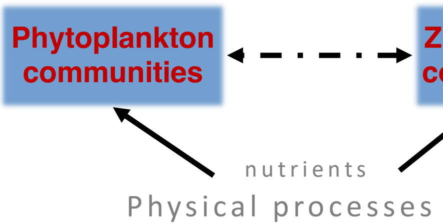
- CNN, ConvLSTM, **U-Net (the best one)**
- Still Chl amplitude underestimation

  → difficulties to reproduce the seasonal and inter-annual extrema

Ongoing work (Mahima Lakra post doc. IMT) :

- Sequence to sequence (forecasting): $\partial_t Y = f(\partial_t X),\ \partial_t Y = f(\partial_t X, \partial_t Y)$....
- architecture « de type operateur neuronaux » , e.g. Fourrier operators (FourCastNet)

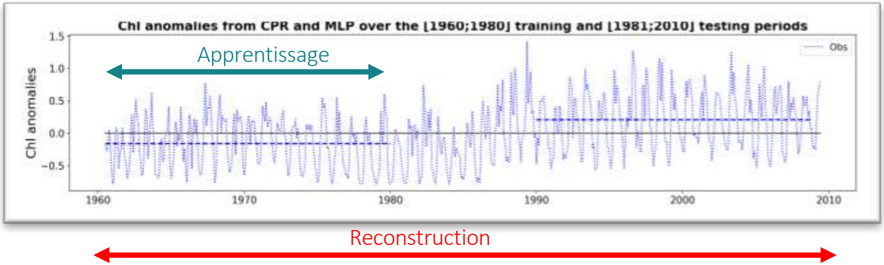
WP2b Elucidating abiotic and phyto-zooplankton inter-specific relationships from *in situ obs.*



In situ obs. in contrasted BGC environment:

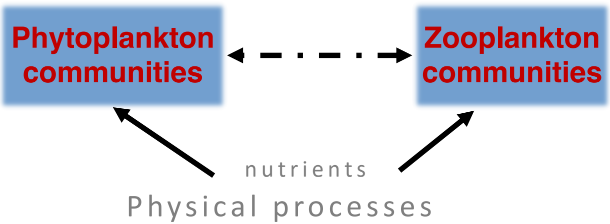
- Oligotrophic area (Hawaii)
- Coastal upwelling (California)
- High latitudes (North Atlantic)

In situ obs.(Mer du Nord)
Continuous Plankton Recorder (CPR)



Aside about reproducing extrema

WP2b Elucidating abiotic and phyto-zooplankton inter-specific relationships from *in situ obs.*

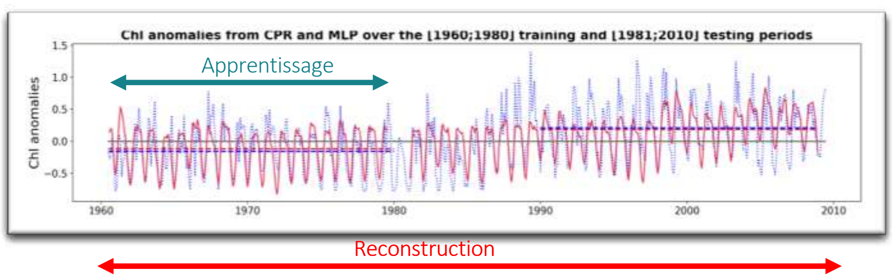


In situ obs. in

contrasted BGC environment:

- Oligotrophic area (Hawaii)
- Coastal upwelling (California)
- High latitudes (North Atlantic)

In situ obs.(Mer du Nord)
Continuous Plankton Recorder (CPR)



Thanks for your attention

Any questions?

E. Martinez, T. Gorgues
E. Pauthenet, M. Roux

M. Lengaigne

R. Fablet, L. Dumetz
M. Lakra

O. Pannekoucke

M. Sourisseau

D. Raitsos

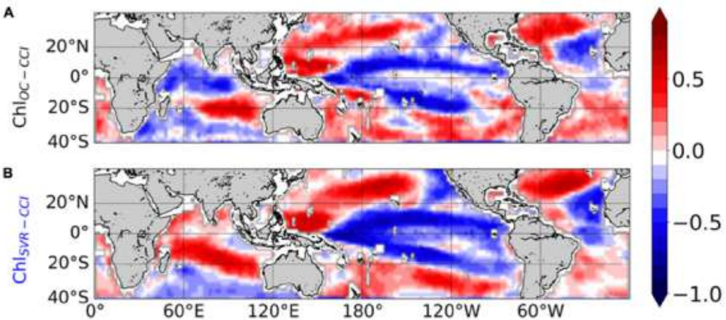
M. Messié

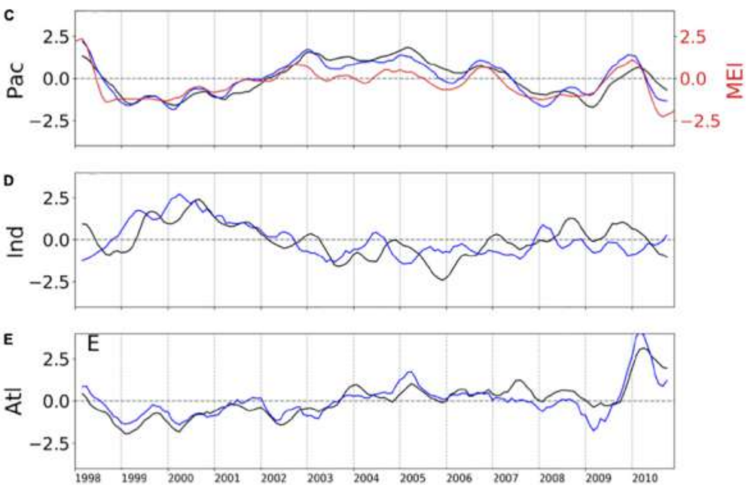
Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	NEMO-PISCES model	1979-2010
SVR	NEMO-PISCES model	Satellite (OC-CCI)	1998-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Training: 7% randomly
Validation: 93%

1st mode of interannual EOF:





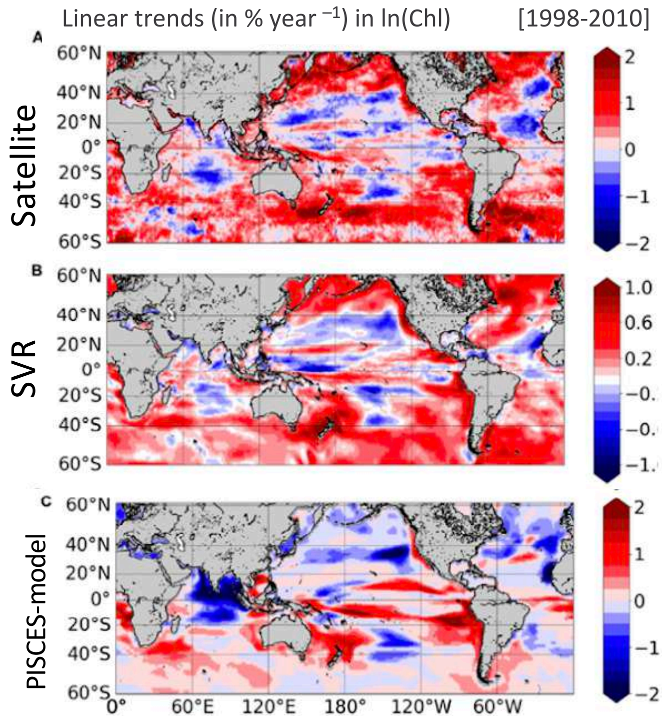
Regional biais

Martinez et al., Frontiers in Marine Sciences 2020

Method	Physical predictors	Chl	Period
SVR	NEMO-PISCES model	Satellite (OC-CCI)	1998-2010

1°x1°grid monthly
Predictors:
7 + lon, lat, time

Training: 7% randomly
Validation: 93%



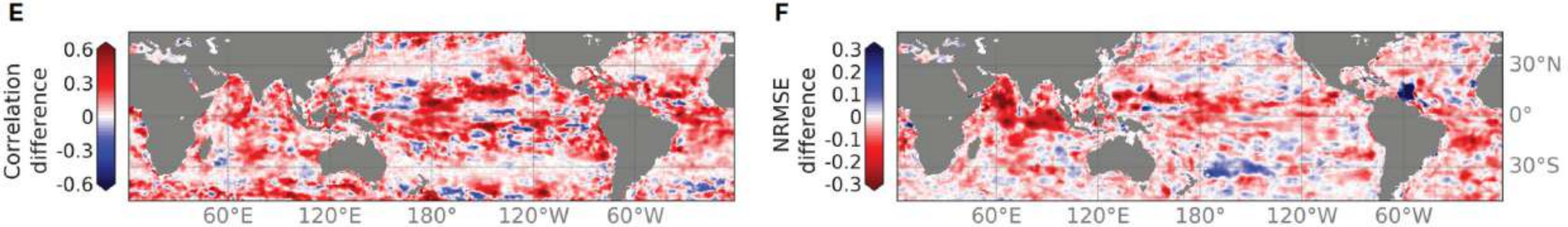
Same physics from the numerical model NEMO

Martinez et al., Frontiers in Marine Sciences 2020

Test 2012-2015:

Predictors		Global scatterplot			Corr. Seas. PC	Corr. Inter. PC	N param	Time computation
	Model	R^2	RMSE	Slope				
9 (+ bathymetry + continental binary mask)	CNN_1	0.84	0.31	0.85	0.99	0.95	99 457	4h54
	CNN_{MM8}	**0.87**	**0.28**	**0.90**	**1.00**	**0.96**	803 920	39h

Difference CNN_{MM8} vs. CNN_1 when compared to satellite Chl



Modes of CNN_{MM8} can regionally learn specific phytoplankton responses to the physical forcing

→ better capture some regional processes than CNN_1

Roussillon et al., Frontiers in Marine Sciences 2023