Rentrée 2019

PROPOSITION DE SUJET DE THESE

Identification du projet

Acronyme du projet *(8 caractères maximum)* **: DREAMS**
Intitulé du projet *en langue française* : Approche par apprentissages profonds pour reconstruire de longues séries chronologiques de la biomasse phytoplanctonique et évaluer les mécanismes dynamiques sous-jacents en lien avec les cycles climatiques.
Intitulé du projet en langue anglaise : Deep leanning approaches to **R**econstruct long time-series of phytoplankton biomass and **E**valuate the underlying dyn**A**mic **M**echanisms related to climate cycle**S**

Établissement porteur du projet : IRD/LOPS
Ecole Doctorale : Ecole Doctorale des Sciences de la Mer et du Littoral (EDSML)

Identification du-de la responsable du projet (futur-e directeur-trice de thèse)

Nom du laboratoire d'accueil : Laboratoire d'Océanographie Physique et Spatiale (LOPS)
Code du laboratoire (U/UMR/USR/EA/JE/…) : UMR 6523
Directeur du Laboratoire : F. Ardhuin
Nom de l'équipe de recherche : Océan et Climat (50%) et SIAM (50%)

Nom et prénom du directeur de thèse (HDR), porteur du projet : Christophe Maes (LOPS°) jusqu'à la soutenance de HDR de Elodie Martinez.

Co-directeur-trice de thèse et co-encadrant scientifique : Elodie Martinez (LOPS)
Contact to apply: elodie.martinez@ird.fr

Présentation du projet (en langue française ou anglaise, 2 à 3 pages)

Résumé du projet *(4000 caractères maxi espaces compris) :*

Le phytoplancton joue un rôle clé dans le cycle du carbone et la chaîne alimentaire marine. La caractérisation des oscillations décennales du phytoplancton en termes de biomasse, de structure des communautés et de production primaire est primordiale car, selon leur phase de régime, les oscillations décennales peuvent accentuer, atténuer ou masquer les tendances séculaires dues au réchauffement climatique. En effet, les incertitudes entourant la variabilité décennale du phytoplancton rendent difficile l'extraction de leurs différents cycles naturels (c.-à-d. saisonniers, interannuels et décennaux) de la tendance anthropogénique. Peu d'études existent à l'échelle mondiale, et la caractérisation des cycles décennaux du phytoplancton ainsi que la compréhension des mécanismes physiques et biogéochimiques sous-jacents restent encore ouvertes. De plus, les changements biologiques observés au cours des phases chaudes des cycles décennaux du climat peuvent donner un aperçu de la façon dont les changements climatiques futurs modifieront le cycle du carbone et les réseaux trophiques marins.

La caractérisation de cette variabilité décennale s'avère une tâche ardue compte tenu de la disponibilité limitée d'observations appropriées pour les échelles de temps considérées (une 20^{aine} d'années pour les observations satellites). Bien que les composantes physiques des modèles physique-biogéochimique réalistes sur de longues séries temporelles (depuis le milieu du 20ème siècle) soient exploités avec une relative confiance, la variabilité des composantes biogéochimiques (et donc phytoplanctonique) de ces modèles i) s'avèrent fortement dépendantes du modèle biogéochimique mis en œuvre et, 2) bien que ces modèles semblent capables de résoudre à un degré toujours meilleur la

variabilité biogéochimique saisonnière ou interannuelle, ils diffèrent néanmoins dans la reproduction des observations décennales. Par conséquent, la stratégie de ce travail de thèse consiste à développer une méthode statistique originale (c.-à-d. une approche par apprentissage profond –Deep learning) qui combine les données radiométriques (dont la couleur de l'océan) et physiques dérivées des observations satellitaires au cours des 20 dernières années avec des modèles physiques océaniques (couvrant une période plus longue que les observations satellites) pour reconstruire des séries temporelles de paramètres biologiques et optiques sur plusieurs décennies, comme la concentration de chlorophylle-a (Chl, une mesure de la biomasse phytoplanctonique). La variabilité aux échelles décennale et globale de la Chl, ainsi que celle d'autres indicateurs concernant la structure des communautés phytoplanctoniques et la production primaire associée, sera caractérisée à partir des séries temporelles reconstruites.

L'objectif principal de cette thèse est de documenter et de comprendre la variabilité décennale passée de la biomasse phytoplanctonique, de leurs assemblages et de la production primaire associée à l'échelle globale, ainsi que les mécanismes sous-jacents.

Phytoplankton plays a key role in carbon cycling and support the ocean food web. Characterizing the decadal oscillations of phytoplankton in terms of biomass, community composition and primary production is crucial as, depending on their regime phase, decadal oscillations can emphasize, weaken or mask the secular climate-related trends. Indeed, uncertainties about decadal variability of phytoplankton make challenging the extraction of their different natural cycles (i.e., seasonal, interannual and decadal) from the anthropogenic trend. Few studies exist at global scale and the characterization of decadal phytoplankton cycles as well as the understanding of the physical and biogeochemical underlying mechanisms still remain open questions. Moreover, the observed changes during positive phases may also provide insight into how future climate-driven change will alter the carbon cycle and the marine food web.

Characterizing this decadal variability proves to be a challenging task considering the limited availability of appropriate observations for the timescales considered (20 years of satellite observations). Realistic physical-biogeochemical models run over long time-series (since the middle of the 20th century) may have been used with a relative confidence for the physical part, but i) phytoplankton variability in those models has been proven to highly depend on the biogeochemical model implemented and, ii) while they seem able to resolve seasonal to interannual biogeochemical variability to an ever-improving degree, yet they diverge in reproducing decadal observations. Consequently, the strategy is, here, to **use an original statistical method (i.e., a deep learning approach) that couples radiometric (ocean color) and physical data derived from satellite observations over the last 20 years with ocean physical models (covering a longer period than satellite observations) to reconstruct multi-decadal time series of biological/optical parameters such as the Chlorophyll-a concentration (Chl, a proxy of phytoplankton biomass). Variability at decadal and global scale of Chl, as well as other indicators about phytoplankton community structures and associated primary production, will then be inferred from their reconstructed time series.**

The main objective of this thesis is to document and understand the past decadal variability of the phytoplankton biomass, assemblages and associated production as well as the underpinning driving mechanisms at global scale.

Présentation détaillée du projet :

1 - Hypothèse et questions posées, identification des points de blocages scientifiques

Phytoplankton—the microalgae that populate the upper lit layers of the ocean—fuels the oceanic food web and regulates oceanic and atmospheric carbon dioxide levels through photosynthetic carbon fixation. Seasonal and inter-annual cycles of phytoplankton biomass are now relatively well characterized at both regional and global scales, thanks to the large amount of studies based on radiometric satellite observations collected since the launch of the SeaWiFs sensor onboard the Seastar spacecraft at the end of 1997 (McClain et al., 2004). While continuous since the end of the 1990's, the time span of this satellite derived ocean color dataset (~20 years) is still too short to thoroughly investigate decadal Chl variations. The built of a consistent time series between the 1979-1983 CZCS dataset and the 1998-2002 SeaWiFS dataset by Antoine et al. (2005) lead on average to an increase of ~20% for the world ocean phytoplankton biomass. These results are almost unique, and only one another similar study was carried out blending satellite Chl with in situ observations and showing on the contrary a slight 6% Chl reduction (Gregg and Conkright, 2002). Studies based on the reprocessed CZCS-SeaWiFs time series by Antoine et al. (2005) allowed to provide a first insight on the impact of large scale decadal climate cycles on global surface Chl (Martinez et al., 2009; D'Ortenzio et al., 2012). Establishing the existence of Chl decadal variations and trends requires long records. Consequently, it is still not possible to distinguish a long-term response to climate change from natural variability in all regions, especially where the signal-to-noise ratio is small. Indeed, *in-situ* biogeochemical observatories provide such a long and continuous time series but the scarcity of the network of such open ocean observatories, which are mainly located in the northern hemisphere and only represents 9-15% of the global ocean surface (Henson et al., 2016), do not allow a global assessment or even a basin-wide assessment of these decadal variations.

The unavailability of global scale observations over a continuous time-series longer than two decades led the scientific community to strongly rely on coupled physical-biogeochemical ocean model simulations to investigate Chl decadal variability. While models are able to resolve seasonal to interannual biogeochemical variability to an ever-improving degree (Dutkiewicz et al., 2001; Wiggert et al., 2006; Aumont et al., 2015), yet they diverge in reproducing decadal observations (Henson et al., 2009a,b; Patara et al., 2011), in particular phytoplankton regime shifts (Henson et al., 2009b). Biological models differ from case to case, which makes it difficult to separate differences in physical responses from differences in biological models. In addition, two of the largest sources of model uncertainties rely on i) the initial nutrient concentrations that contribute on the biogeochemical and biological parameterizations (Laukfotter 2015) and ii) the response of physiological processes of organisms to changes in their environment (e.g., temperature) which remains uncertain and thus difficult to model and predict (Sarmiento et al., 2004).

Nevertheless decadal cycles of phytoplankton (in terms of chlorophyll biomass, community composition and carbon fluxes) have to be well characterized for three major reasons: (1) They can accentuate, weaken or even mask the secular (climate-related) trends (the recent debate about the observed cooling in the context of climate change illustrates the crucial need for better understanding decadal variability); (2) The observed changes in phytoplankton during warm phases of decadal cycles may provide insight into how future climate warming-induced changes will alter carbon cycle and marine food web; (3) Predicting future global changes in oceanic phytoplankton with some confidence requires first a comprehensive understanding of how climate changes influenced them in the past.

Therefore, characterizing biogeochemical decadal cycles is not only crucial in the context of climate change, it is also definitely challenging!

This thesis is expected to provide answer to the following specific questions:

- **How did the different climate modes modulated the phytoplankton response over the last decades? (i.e., seasonal, interannual and decadal modes)**
- **Which subsequent changes in phytoplankton community structure and associated PP occurred?**
- **What are the physical mechanisms (such as vertical mixing, advection…) underlying the observed biogeochemical changes?**

2 - Approche méthodologique et techniques envisagées :

To answer these questions and fully apprehend phytoplankton and oceanic dynamics at the decadal scale, observations have to be available over sufficiently long time series at the basin/global scale (i.e., through remote sensing). The underlying idea is that deep learning approaches can allow the reconstruction of long-term time series of Chl and particulate backscattering coefficient (the ratio of these two parameters provide a more accurate insights on phytoplankton biomass changes than Chl only, Behrenfeld et al., 2005) and phytoplankton community structure (Uitz et al., 2010 ; Deboissieu et al., 2015) based on radiometric satellite observations *vs*. physical oceanic and atmospheric metrics. To do so, it is considered that the distribution of phytoplankton is i) to a large extent controlled by physical process (i.e., bottom up process which drive nutrient availability toward the upper lit layer of the ocean) and ii) the impact of predation by zooplankton (i.e., top-down controm) which can also be influenced by their physical environnement, specifically temperature (Beaugrand et al., 2002) and thus can be implicitly taken into account by the physical forcing. While statistical reconstructions are increasingly applied to extend physical metrics back in time, past reconstruction of surface Chl has only been performed, to our knowledge, by Schollaert et al. (2017) who focused on the tropical Pacific thanks to a linear canonical correlation analysis. Several steps would be considered:

- First step: Reconstruct multi-decadal time series of phytoplankton biomass and phytoplankton community structure at global scale.
- Second step: Determine the phytoplankton response to the decadal climate modes.
- Third step: Identify the physical mechanisms underlying the biological activity.

At the beginning of the thesis, the reliability of several Deep-Learning methods will be assessed to reconstruct multi-decadal time series of phytoplankton biomass and Phytoplankton Community Structure (PCS) at global scale.

From a methodological point of view, geophysically-sound learning schemes will be of key interest, especially residual neural networks which can be regarded as numerical schemes for Ordinary and Partial Differential Equations (ODE and PDES espectiveley) (He et al., 2016; Fablet et al., 2018). Biases and robustness of the different methods will be assessed, specifically to reconstruct the low frequency signal related with climate cycles (step 1). Within the framework of the PHYTODEV CNES-project 2016-2018 (PI E. Martinez), the feasibility and robustness of Support Vector machine non-linear Regressions (a basic machine learning method less sophisticated than Deep-Learning) have been assessed on a synthetic data set derived from a coupled bio-physical global ocean model (article in prep, but see for example Jouini et al., 2013). Several Chl-reconstruction time-series have been performed using a small amount of predictors (e.g., 7). These preliminary informations will help evaluating the improvement provided by Deep Learning mehods, thanks in particular to the present support/collaboration of the IMT and Ifremer members (see below section 9/ Contexte scientifique et partenarial).

Thanks to the reconstructed multidecadal phytoplankton biomass time-series and PCS, the phytoplankton response to decadal climate modes could be determined. More specifically, which changes occurring in the magnitude and timing of phytoplankton seasonal and interannual cycles depending on the decadal warm and cold phases, as well as in phytoplankton community structure and primary production could be determined (step 2).

Finally, sensitivity tests performed on the different physical metrics used as predictors will help i) to evaluate their relative contributions to biological changes across oceanic regions and biogeochemical provinces, ii) to identify the physical mechanisms (i.e., vertical mixing/advection variations…) underlying the biological activity. Finally, Deep Learning is not so much a black box anymore. According to the method used, behaviors are determined by their weights and biases, therefore we will also take adavantage of the ability of learn differential representations from data to explore the identification of governing equations for the processes of interest.

3 - Contexte scientifique et partenarial : éléments généraux

The ambitious interdisciplinary approach proposed here is based on the strong expertise of the various collaborators on this thesis topic and their complementary research fields: ocean dynamics and biogeochemistry of the ocean via satellite observations and numerical modelling combined with statistical and artificial intelligence approaches.

- Locally/regionally:
  - **IRD/LOPS**. Elodie Martinez and Thomas Gorgues : physicist-biogeochemist oceanographers, specialists in physical-biogeochemical interactions in response to climate cycles through satellite remote sensing and numerical modelling.
  - **Ifremer Brest Océan et Climat**. Guillaume Maze: physicist oceanographer and statistician, specialist in ocean dynamics related to climate cycles via observations.
  - **IMT-Atlantique/ UMR 6285 LabSTICC**. Ronan Fablet, Pierre Tandeo, Lucas Drumetz: specialists in applied mathematics (statistics, signal processing) and artificial intelligence related to geophysical applications and ocean remote sensing data.

- Nationally:
  - **CNRS/LOV**. Julia Uitz: océanographe biogeochemical oceanographer, specialist in phytoplankton community structures and primary production via satellite remote sensing.

- Internationally:
  - **Georgia Institute of Technology, USA**. Emanuele Di Lorenzo: physicist oceanographer, specialist in climate-marine ecosystem interactions.

Le – la candidat.e (profil souhaité, compétences scientifiques et techniques requises)

This subject is intended for a motivated candidate with a Master's degree in oceanography (e.g. marine biogeochemistry, physics or chemistry of the marine environment) or in information and telecommunication technology (e.g. bigs data, signal processing, artificial intelligence). Strong digital programming skills are required. In particular, a knowledge of programming languages (Fortran, Python, Matlab) and the UNIX/Linux environment is essential. Good communication skills in English will be appreciated.

Références bibliographiques

Antoine, D., A. Morel, H. R. Gordon, V. F. Banzon, R. H. Evans (2005). Bridging ocean color observations of the 1980s and 2000s in search of long-term trends, J. Geophys. Res., 110, C06009.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8, 2465–2513, 2015, doi:10.5194/gmd-8-2465-2015

Beaugrand G, Reid PC, Ibanez F, Lindley JA, Edwards M (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. Science, 296, 1692–1694.

Behrenfeld, M. J., Boss, E., Siegel, D. A. and Shea, D. M.: Carbon-based ocean productivity and phytoplankton physiology from space, Global Biogeochem. Cycles, 19(1), GB1006, doi:10.1029/2004GB002299, 2005.

De Boissieu, F., C. Menkes, C. Dupouy, M. Rodier, S. Bonnet, M. Mangeas, R. J. Frouin (2014). Phytoplankton global mapping from space with a Support Vector Machine algorithm, International Society for Optics and Photonics, p. 92611R–92611R.

D'Ortenzio, F. D. Antoine, E. Martinez, M. Ribera d'Alcalà (2012). Phenological changes of oceanic phytoplankton in the 1980s and 2000s as revealed by ocean-color remote-sensing observations, Global Biogeochem. Cy.,, Vol. 26, GB4003.

Dutkiewicz, S., M. Follows, J. Marshall, W.W Gregg (2001). Interannual variability of phytoplankton abundances in the North Atlantic, Deep Sea Res. II, Volume 48, Pages 2323-2344.

Ronan Fablet, R., S. Ouala, C. Herzet (2018). Bilinear residual Neural Network for the identification and forecasting of dynamical systems. EUSIPCO 2018, European Signal Processing Conference, Sep 2018, Rome, Italy.

Gregg, W. W., and M. E. Conkright, (2002). Decadal changes in global ocean chlorophyll. Geophysical Research Letters 29, 1730.

He, K., X. Zhang, S. Ren, J. Sun (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Henson, S. A., J. P. Dunne, J. L. Sarmiento (2009a). Decadal variability in North Atlantic phytoplankton blooms, J. Geophys. Res., 114, C04013.

Henson SA, Raitsos D, Dunne JP, McQuatters-Gollop A (2009b). Decadal variability in biogeochemical models: Comparison with a 50-year ocean colour dataset. Geophysical Research Letters, 36, L21061.

Henson, S. A., C. Beaulieu, and R. Lampitt, (2016), Observing climate change trends in ocean biogeochemistry: when and where. Glob Change Biol, 22: 1561-1571. doi:10.1111/gcb.13152

Jouini, M., M. Lévy, M. Crépon, S. Thiria (2013). Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method, Remote Sens.Environ., 131, 232-246.

Laufkötter, C., M. Vogt, N. Gruber, M. Aita-Noguchi, O. Aumont, L. Bopp et al. (2015). Drivers and uncertainties of future global marine primary production in marine ecosystem models, Biogeosciences, 12, 6955-6984, https://doi.org/10.5194/bg-12-6955-2015.

Martinez, E., D. Antoine, F. D'Ortenzio, B. Gentili (2009). Climate-driven basin-scale decadal oscillations of oceanic phytoplankton, Science, 36, 1253-1256.

McClain, C. R., G. Feldman, and S. Hooker (2004). An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series, Deep Sea Res. Part II, 51, 5–42, doi:10.1016/j.dsr2.2003.11.001.

Patara, L., M. Visbeck, S. Masina, G. Krahmann, M. Vichi (2011). Marine biogeochemical responses to the North Atlantic Oscillation in a coupled climate model, J. Geophys. Res., 116, C07023.

Sarmiento, J. L., R. Slater, R. Barber, L. Bopp, S. Doney, et al. (2004). Response of ocean ecosystems to climate warming, Global Biogeochem. Cy., 18, GB3003.

Schollaert Uz, S., A. J. Busalacchi, T. M. Smith, et al. (2017). Interannual and Decadal Variability in Tropical Pacific Chlorophyll from a Statistical Reconstruction: 1958–2008, J. Clim. 30, 7293–7315. doi:10.1175/JCLI-D-16-0202.1.

Uitz, J., H. Claustre, B. Gentili, D. Stramski (2010). Phytoplankton classspecific primary production in the world's oceans: Seasonal and inter-annual variability from satellite observations. Global Biogeochem. Cy., Vol. 24, No. 3, GB3016.

Wiggert, J. D. , R. G. Murtugudde, J. R. Christian (2006). Annual ecosystem variability in the tropical Indian Ocean: Results of a coupled bio-physical ocean general circulation model, Deep Sea Res. II, Volume 53, Issues 5–7, Pages 644-676.