

Etude de cas : Caractérisation de l'écosystème phytoplanctonique des zones d'upwelling

Hadrien Boulanger, Adam Aalah,
Master TRIED

Abstract—Le phénomène d'upwelling sur la côte Sénégal-Mauritanienne, comme on va le voir, peut être cartographié géographiquement, au le sens où on peut découper la zone en secteurs où son activité diffère, où il est plus ou moins fort tout au moins. Nous tenterons ici de voir si ces secteurs contiennent des écosystèmes phytoplanctoniques caractéristiques en recoupant les classifications géographiques issues de l'analyse de l'upwelling et de l'analyse de différentes espèces de phytoplanctons.

I. INTRODUCTION

Le phénomène de remontées d'eau (ou upwelling) observé au large de la côte Mauritano-Sénégalaise a pu être identifié géographiquement et temporellement à l'aide de l'analyse de données satellites sur 2 variables : la température de surface des eaux, et la concentration en Chlorophylle-a, un pigment mettant en évidence la présence de phytoplancton. En effet, lorsque le phénomène est actif dans une région, la concentration de Chlorophylle-a est significativement plus élevée et la température plus faible [1]. La zone d'upwelling se déplace au cours des saisons et est beaucoup plus large autour de février-mars qu'à son minimum qu'elle atteint à la fin de l'été, puis elle s'élargit de nouveau et continue son cycle annuel. Mais il existe une diversité de phytoplanctons, et on dispose de données sur d'autres pigments qui nous permettent d'étudier la variabilité de ce phytoplancton [2]. Nous tenterons donc de répondre au problème "Y'a-t-il un écosystème phytoplanctonique caractéristique des zones d'upwelling dans la région de la côte sénégal-mauritanienne ?"

II. PRÉSENTATION DES VARIABLES

Pour les variables, nous pouvons exploiter des données satellites sur la température et sur les concentrations en certains pigments présents dans cette région de l'océan. On a de nombreuses informations manquantes. C'est pourquoi lorsque nous faisons une représentation d'une variable sur une période de temps nous moyennons les pixels à chaque localisation sur la période de temps étudiée en ignorant les pixels manquants dans le calcul de celle-ci. Par exemple si on fait une moyenne de SST sur l'année 2004 (365 jours) et qu'il y a eu 30 jours sans données pour un pixel donné, on affichera la moyenne des 335 jours de données pour ce pixel.

A. Analyse univariée

On observe les moyennes sur l'année pour se donner une idée globale des variables liées à l'upwelling :

On voit que dans la région, les eaux de surface sont globalement plus chaudes au sud qu'au nord.

SST :

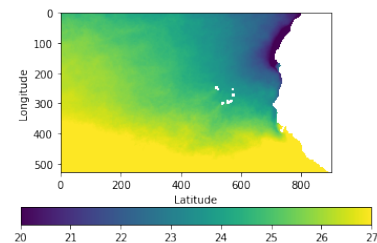


Figure 1. SST moyenne 2004

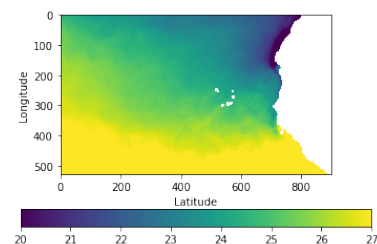


Figure 2. SST moyenne 2005

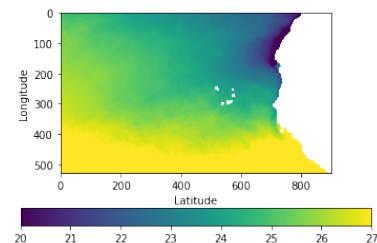


Figure 3. SST moyenne 2006

On voit aussi une concentration de chlorophylle-a bien plus importante au niveau des côtes. Chl-a :

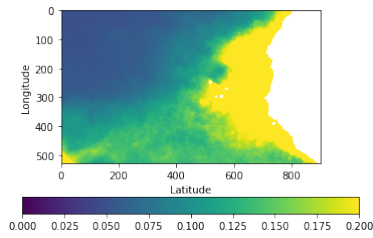


Figure 4. Chla moyenne 2004

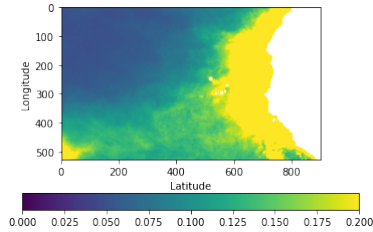


Figure 5. Chla moyenne 2005

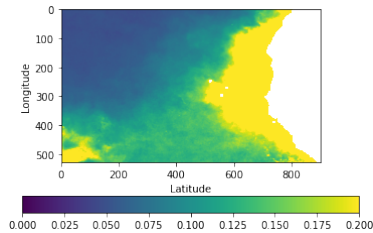


Figure 6. Chla moyenne 2006

En faisant des moyennes mensuelles, on peut tracer les séries temporelles liées à toutes les variables (voir annexe).

B. Analyse multivariée

L'analyse multivariée montre (voir annexe) que les variables les plus corrélées entre elles sont les pigments chla et chlb. La réalisation d'une ACP sur les variables sst, PAR, ZEU, Chla, Chlb, Zea, 19HF, Fuco, nous permet de voir quelles variables contribuent le plus aux 3 premiers axes de l'ACP, et lesquelles évoluent ensemble :

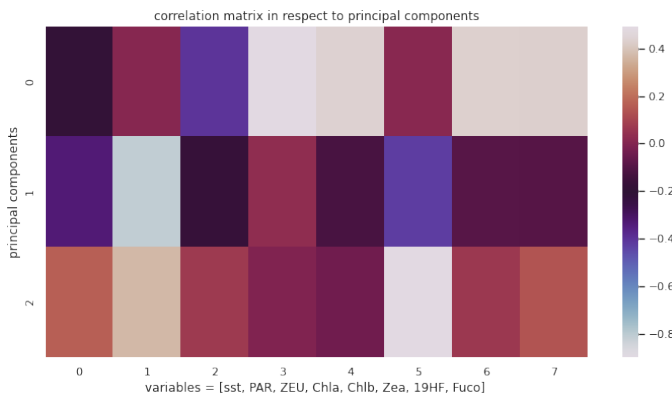


Figure 7. Contribution sur les axes de l'ACP

On voit que globalement les pigments Chla, Chlb, 19HF et Fuco évoluent ensemble sur les axes principaux de l'ACP : ont une contribution positive sur l'axe 0 et presque nulle sur l'axe 1 et sur l'axe 2. Zea et SST évoluent différemment et ne contribuent quasiment pas à l'axe 0 alors qu'ils sont légèrement négatifs sur l'axe 1, et contribuent positivement à l'axe 2. Voici la projection des individus sur le plan des axes 1 et 2 de l'ACP.

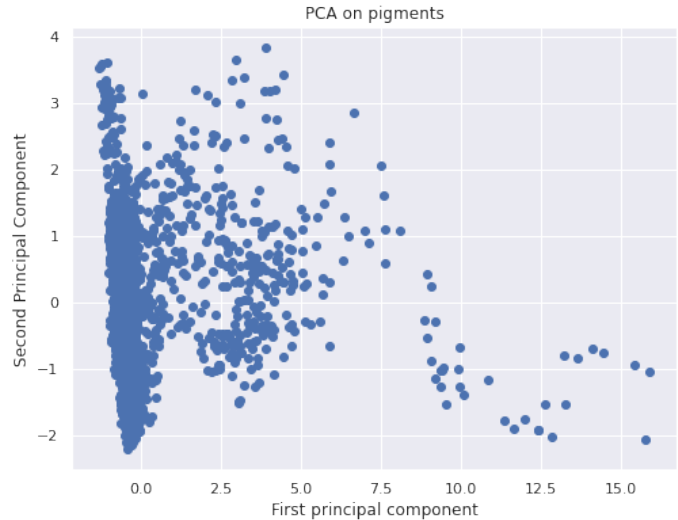


Figure 8. ACP projection plan 1

Pour les clusterings servant à l'identification des zones d'Upwelling, on utilisera les variables utilisées par Farikou et al.[1] : SST (température de surface) et Chl-a (concentration de Chlorophylle-a en mg.m-3).

Dans les clusterings servant à cartographier la variabilité phytoplanctonique nous utiliserons donc des variables utilisées par R. El Hourany et al.[2] : également la variable Chl-a avec les variables de concentrations: Chlb (Chlorophylle-b en mg.m-3), Fuco (Fucoxanthine mg.m-3), 19HF (19-Hexanoyloxy-fucoxanthine en mg.m-3) et Zea (Zeaxanthine en mg.m-3). Nous disposons sur ces variables de données quotidiennes d'images (901*531 pixels, 1 pixel représente 4km) de la côte Sénégal-Mauritaniennes sur 3 ans : du 1er janvier 2004 au 22 décembre 2006.

III. MÉTHODE

Pour répondre à la problématique nous souhaitons nous appuyer sur les travaux de clustering de O. Farikou et al.[1] pour identifier saison par saison les zones où l'upwelling est présent, et dans le même temps étudier la variabilité du phytoplancton saison par saison avec d'autres clustering grâce aux méthodes de R. El Hourany et al.[2]. Ainsi, en calculant les similarités entre les clusterings nous pourrions voir s'il existe des écosystèmes phytoplanctoniques caractéristiques aux zones d'Upwelling. Ces travaux de clustering feront intervenir des cartes auto-organisatrices (SOM).

COMMENT TRAITER LE BRUIT ?

Par exemple, on peut observer pour le mois de février 2004 la concentration de chlorophylle-b.

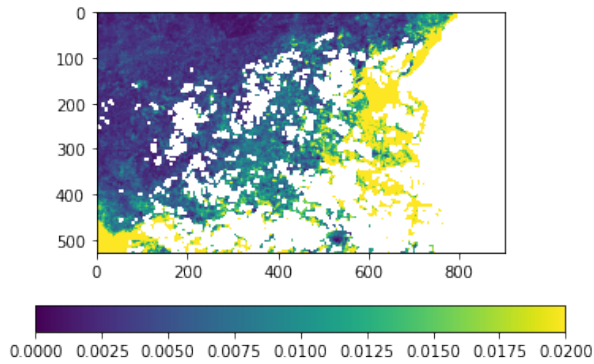


Figure 9. Chlb moyenne février 2004

Cette fois les données sont très bruitées, car de nombreux pixels sont orphelins de données pendant les 29 jours du mois de février 2004. Cela peut être dû aux nombreux nuages qui passent au-dessus de ces zones. On peut améliorer cela en combinant les mois de février des 3 ans de l'étude, car notre étude sera surtout saisonnière étant donné le caractère saisonnier du phénomène d'upwelling. On a donc la moyenne (nanmean) des données sur 3 mois. Il n'y a presque plus de données manquantes.

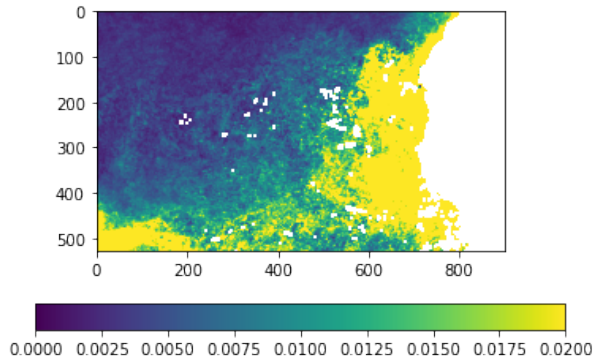


Figure 10. Chlb moyenne février

La figure précédente montre ainsi la concentration moyenne en chlorophylle-b au cours des mois de février des années 2004, 2005 et 2006, soit, la concentration moyenne en chlorophylle-b sur les 3 périodes de plus fortes upwelling. Désormais, nous travaillerons donc sur 3 périodes de 4 mois : Janvier-Avril des 3 années, Mai-Août des 3 années, Septembre-Décembre des 3 années. Ainsi il n'y aura plus de données manquantes

POOLING

Pour réduire les temps de calcul, nous effectuerons un pooling de 3x3 pixels sur toutes les images, qui deviennent donc des images de taille 177*300

IV. PLAN

Grâce aux techniques de visualisations de pooling et de débruitage on a pu traiter les données et ainsi on pourra procéder à des analyses des différentes variables avant de pouvoir faire nos clusterings :

- 1) le clustering de Janvier à Avril des zones d'upwelling (var SST et CHLA)
- 2) le clustering de Janvier à Avril de la diversité du phytoplancton (var CHLA, CHLB, FUCO, ZEA, HF19)
- 3) le clustering de Mai à Août des zones d'upwelling (var SST et CHLA)
- 4) le clustering de Mai à Août de la diversité du phytoplancton (var CHLA, CHLB, FUCO, ZEA, HF19)
- 5) le clustering de Septembre à Décembre des zones d'upwelling (var SST et CHLA)
- 6) le clustering de Septembre à Décembre de la diversité du phytoplancton (var CHLA, CHLB, FUCO, ZEA, HF19)

Ainsi, on pourra étudier la variabilité saisonnière de la diversité du phytoplancton et on pourra mettre en évidence les écosystème phytoplanctoniques caractéristiques aux zones d'upwelling.

V. PREPARATION DES DONNÉES

Nous allons donc travailler pour nos clusterings sur les données journalières de nos 2 périodes (février-mars et août des 3 années). Les "nan" des données de chaque jour seront remplacés par la moyenne sur la période. Par exemple, voici à quoi ressemble la donnée Chla pour le 1er février 2004 avant traitement :

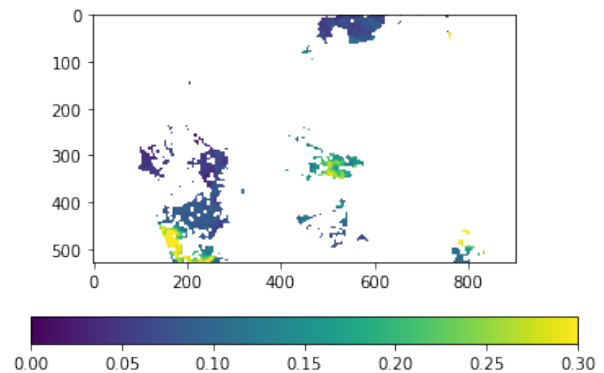


Figure 11. Chla 1er février 2004

Après application de la fonction Nanmean pour toutes les données de février et mars des 3 ans (soit 6 mois de données, suffisant pour avoir au moins une donnée pour chaque pixel)

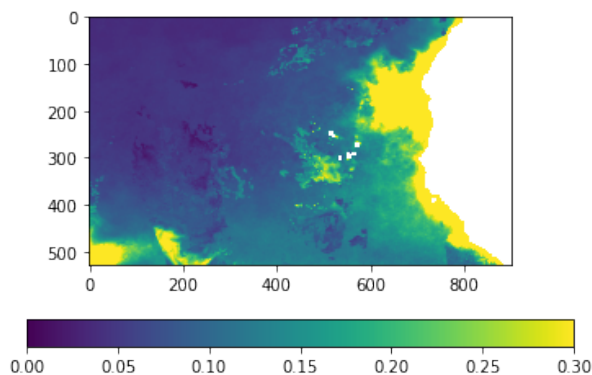


Figure 12. Chla 1er février 2004/moyenne de février-mars

Après ce traitement effectué on travaille donc sur un tableau dont les lignes sont l'ensemble des pixels (53100 après pooling de taille 3x3), et les colonnes l'ensemble des jours de la période étudiée, ici 4 mois. Sur chaque période Janvier-Avril, Mai-Aout et Septembre-Décembre on effectuera 2 clusterings :

	SST	Chla
Pixel1	SST pixel1	Chla pixel1
Pixel2		
...		
Pixel53100	SST pixel53100	Chla pixel53100

Figure 13. Tableau de données liées à l'Upwelling

	Chla	Chlb	FuCo	HF19	Zea
Pixel1	Chla pixel 1	Chlb pixel 1	FuCo pixel 1	HF19 pixel 1	Zea pixel1
Pixel53100	Chla pixel 53100	Chlb pixel 53100	FuCo pixel 53100	HF19 pixel 53100	Zea pixel53100

Figure 14. Tableau de données liées au phytoplancton

VI. CLUSTERING

Tous les clusterings suivants ont été réalisés à l'aide de cartes auto-organisatrices (SOM) de Kohonen, de $20 \times 20 = 400$ neurones, puis de CAH pour la classification des neurones. L'ensemble de ces clusterings est réalisé pour 5 classes, nombre de classe qui à chaque fois semblait le plus pertinent en observant les dendrogrammes des CAH sur les neurones des différentes cartes (voir annexe).

A. Clustering Janvier-Avril

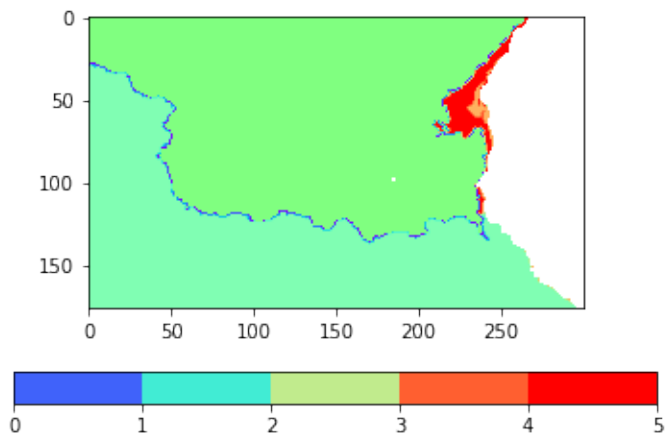


Figure 15. clustering entre janvier et avril sur SST et CHLA

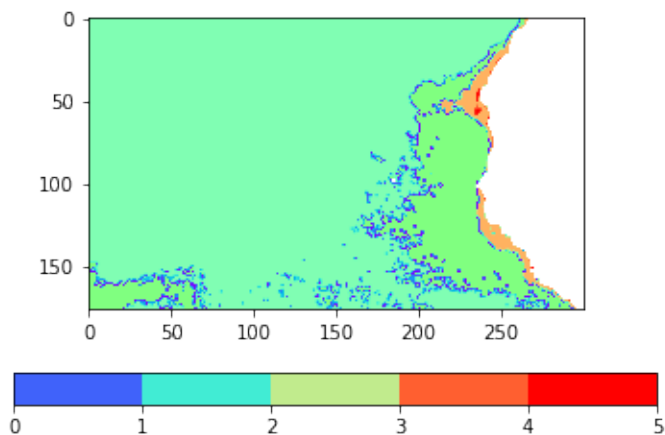


Figure 16. clustering entre janvier et avril sur Chla, Chlb, Fuco, 19HF, Zea

La similarité de Rand entre les 2 clustering précédents est 11.32%

B. Clustering Mai-Août

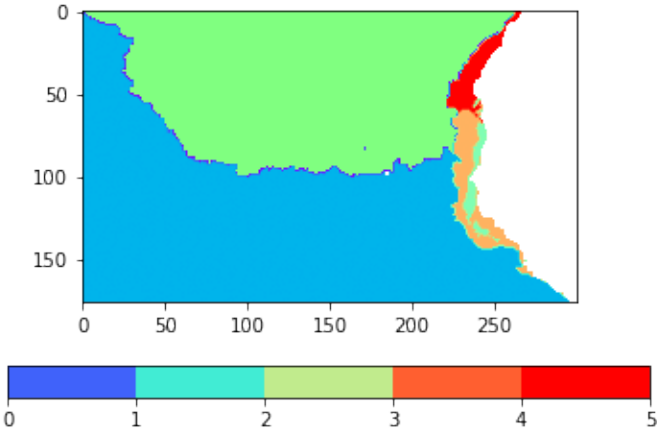


Figure 17. clustering entre mai et août sur SST et CHLA

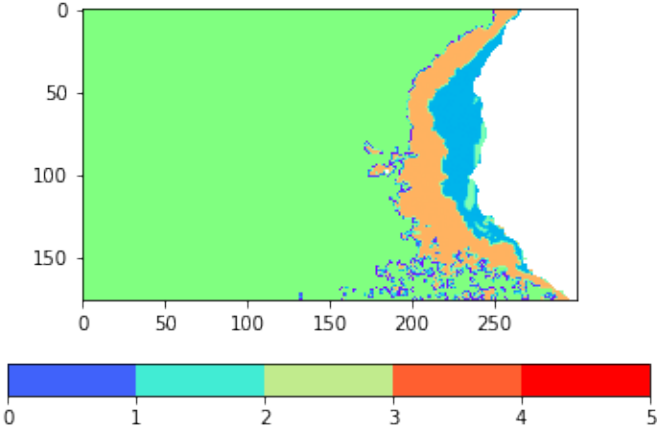


Figure 18. clustering entre mai et août sur Chla, Chlb, Fuco, 19HF, Zea

La similarité de Rand entre les 2 clustering précédents est 12.03%

C. Clustering Septembre-Décembre

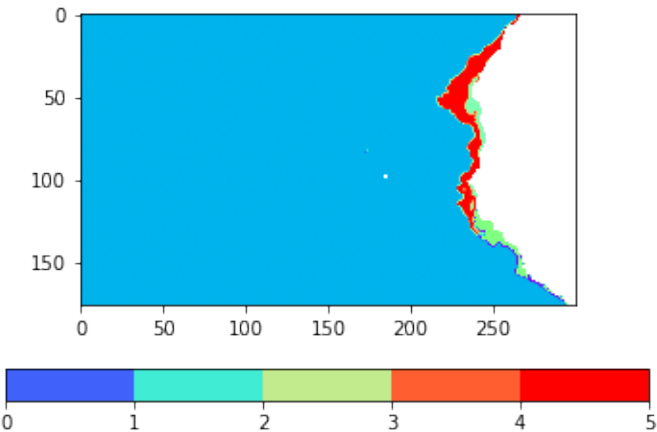


Figure 19. clustering entre mai et août sur SST et CHLA

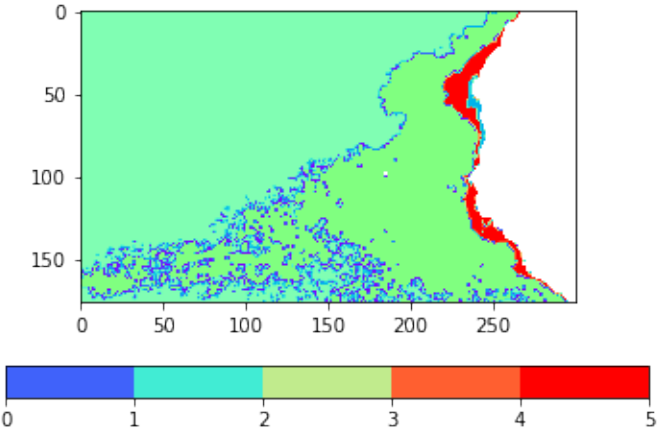


Figure 20. clustering entre mai et août sur Chla, Chlb, Fuco, 19HF, Zea

La similarité de Rand entre les 2 clustering précédents est 11.85%

VII. RESULTATS

Sur chaque période, on voit que la similarité de Rand entre le clustering des zones d'Upwelling et la variabilité phytoplanctonique est autour de 12%. Cherchons plus loin :

Upwelling Janv/Avr	[0.11317159, 0.06987615, 0.14288498]	
Upwelling Mai/Août	[0.10445379, 0.12034619, 0.17922055]	
Upwelling Sept/Déc	[0.24992636, 0.33447058, 0.11847802]	
Phyto Janv/Avr	Phyto Mai/Août	Phyto Sept/Déc

Figure 21. Tableau de similarités de Rand

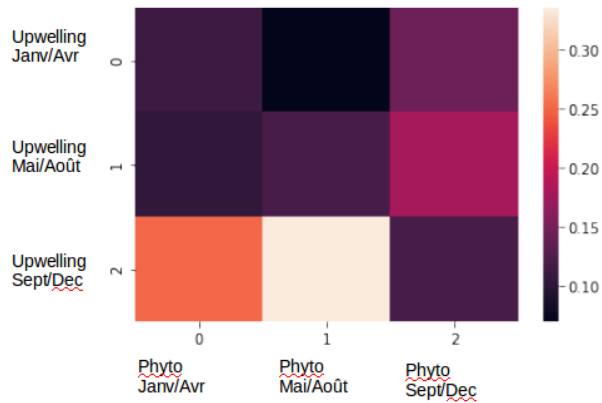


Figure 22. HeatMap des similarités de Rand

Il semble que les 2 phénomènes soient finalement assez peu corrélés.

VIII. CONCLUSION

On a réussi à traiter le problème de manière à éliminer les données manquantes, en travaillant sur des périodes de temps de 4 mois et en remplaçant les données manquantes par la moyenne sur les 4 mois. Les CAH de tous les clusterings montraient de manière indépendante qu'il était toujours plus intéressant de travailler avec un nombre de classe égal à 5. En revanche on n'a pas pu montrer que la similarité entre les phénomènes d'upwelling et la variabilité phytoplanctonique était significative. Il ressort que cette similarité est toujours autour de 12%. Si en effet les 2 phénomènes évoluent au cours du temps en fonction de variables communes (chl_a), on ne peut pas dire qu'il existe réellement des zones géographiques caractéristiques de l'upwelling où est présent un écosystème de phytoplanctons particulier.

REFERENCES

- [1] O. Farikou, S. Sawadogo, A. Niang, J. Brajard, C. Mejia, M. Crepon and S. Thiria, "Multivariate analysis of the senegalo-mauritanian area by merging satellite remote sensing ocean color and sst observations.", 2013.
- [2] Roy El Hourany, Marie Abboud-Abi Saab, Ghaleb Faour, Carlos Mejia1, Michel Crépon, and Sylvie Thiria, "Phytoplankton diversity in the mediterranean sea from satellite data using self-organizing maps.", 2019.

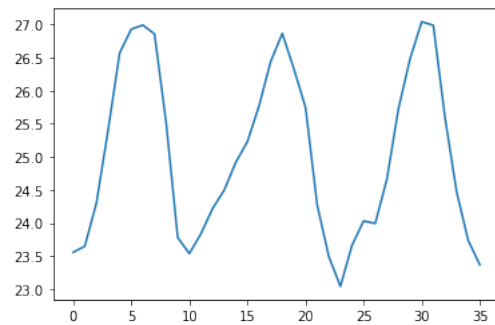


Figure 23. Moyenne mensuelle de SST entre 2004 et 2006

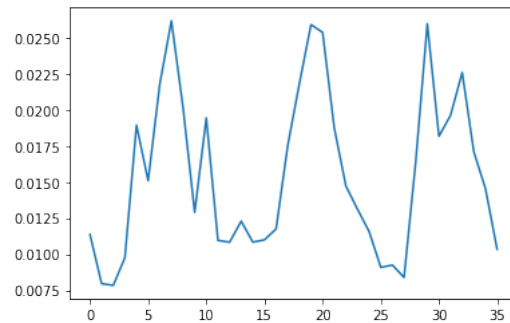


Figure 24. Moyenne mensuelle de Fuco entre 2004 et 2006

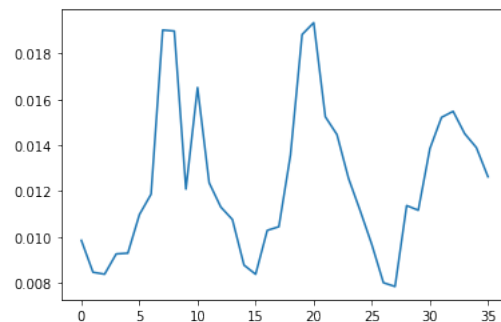


Figure 25. Moyenne mensuelle de Chl-b entre 2004 et 2006

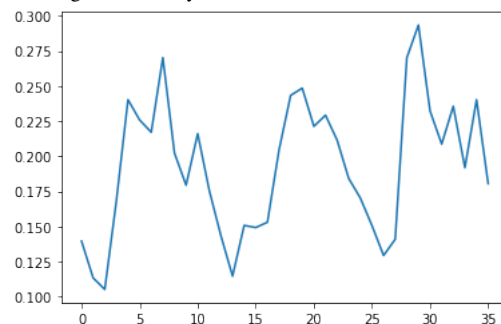


Figure 26. Moyenne mensuelle de Chl-a entre 2004 et 2006

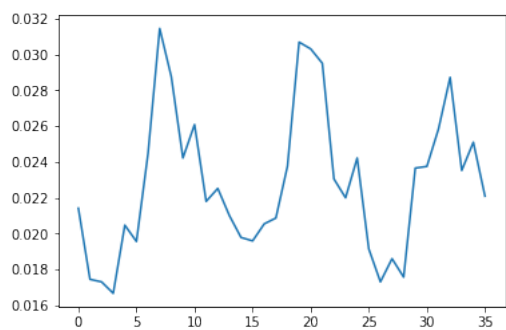


Figure 27. Moyenne mensuelle de 19HF entre 2004 et 2006

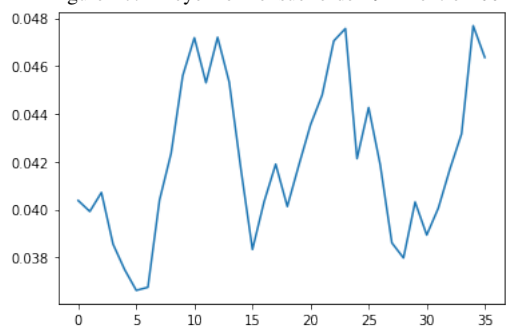
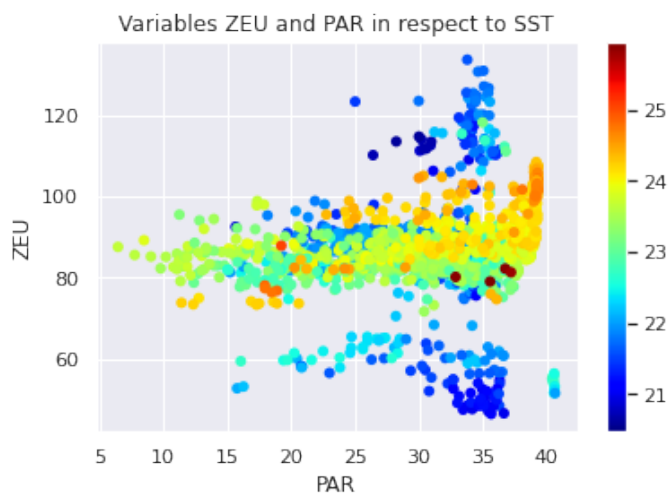
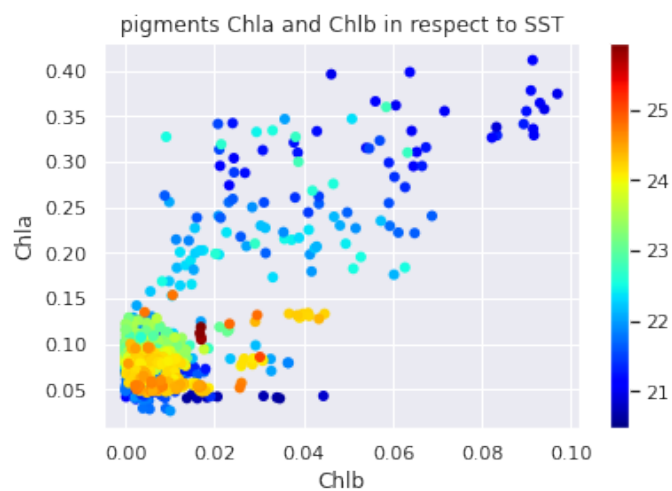
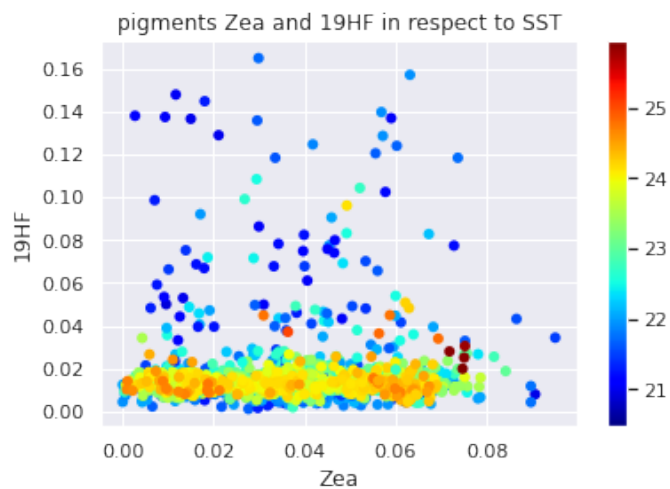


Figure 28. Moyenne mensuelle de Zea entre 2004 et 2006



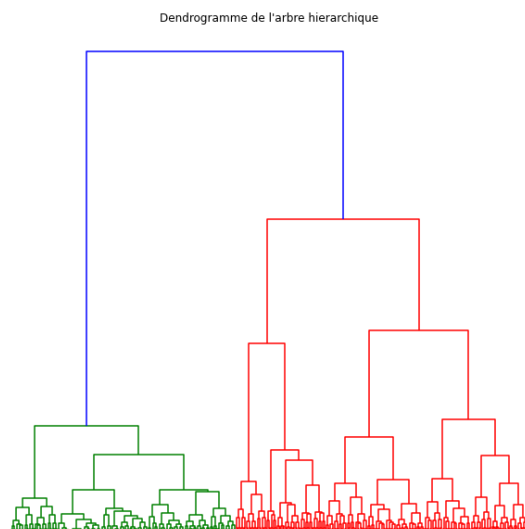


Figure 29. CAH sur les neurones obtenue pour toutes les SOM (même dendrogramme pour toutes les cartes effectuées lors du projet)

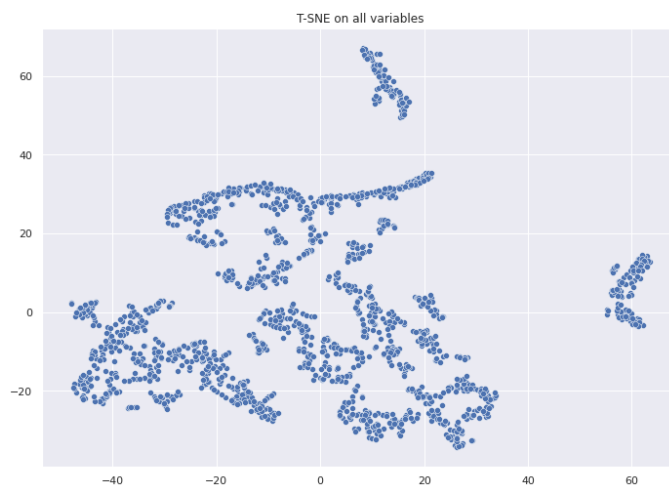


Figure 31. TSNE sur les pigments

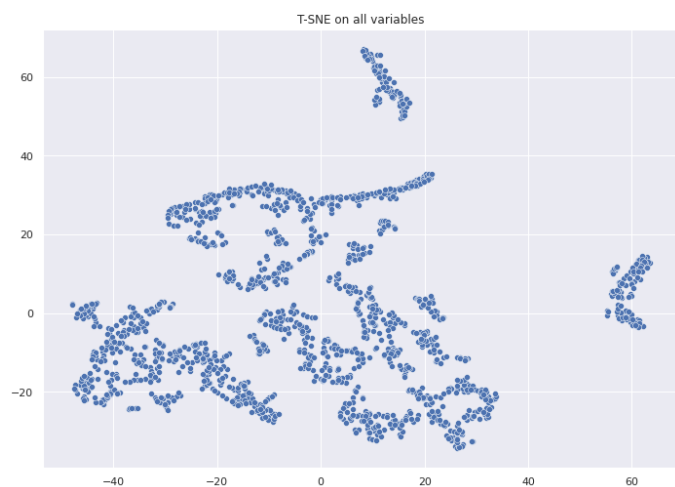


Figure 30. TSNE sur toutes les variables