

AIX-MARSEILLE UNIVERSITÉ
Institut OSU-Pythéas
Institut Méditerranéen d'Océanologie - MIO

MASTER SCIENCES DE LA MER
Parcours Océanographie Biologique et Écologie Marine

**Caractérisation de la dynamique des assemblages
phytoplanctoniques et de leur environnement pélagique à
partir de la série temporelle SOMLIT : approche par
statistique fonctionnelle**

Couteyen Carpaye, Mathilde

Étude réalisée au sein de l'Institut Méditerranéen d'Océanologie

Sous la direction de M. Gérald Grégori et M. David Nérini

Année universitaire

2021-2022

REMERCIEMENTS

Je tiens tout d'abord à remercier mes encadrants Monsieur Gérard Grégori et Monsieur David Nérini pour leurs précieux conseils et le soutien apporté avec bienveillance tout au long de ce stage.

Je tiens à remercier également l'Office Français de la Biodiversité et l'Infrastructure de Recherche ILICO qui ont permis la réalisation de ce stage.

Enfin, je tiens à remercier Monsieur Samuel Couteyen pour ses conseils avisés.

SOMMAIRE

1	Introduction	1
2	Matériel & Méthodes	4
2.1	SOMLIT (Service d'Observation en Milieu LITtoral)	4
2.1.1	Les stations	4
2.1.2	Prélèvements et protocoles	5
2.1.3	Jeux de données	6
2.2	Analyse de séries temporelles	10
2.2.1	Lissage non-paramétrique	10
2.2.2	Saisonnalité	11
2.3	Analyse en Composantes Principales Fonctionnelle	12
2.3.1	Lissage paramétrique	12
2.3.2	ACP fonctionnelle	13
2.3.3	Perturbation de la moyenne	13
2.4	Analyse Factorielle Discriminante	13
2.5	Analyse Canonique des Corrélations	14
3	Résultats	15
3.1	Variations dans le temps	15
3.1.1	Abondance des groupes phytoplanctoniques	15
3.1.2	Température	17
3.2	Variations dans l'espace	20
3.2.1	Abondance des groupes phytoplanctoniques	20

3.2.2	Diffusion lumineuse des groupes phytoplanctoniques	22
3.2.3	Nutriments	24
3.3	Co-variations	25
3.3.1	Abondance et Nutriments	26
3.3.2	Diffusion lumineuse et Nutriments	27
4	Discussion	29
5	Conclusion	33
	Bibliographie	33
	Annexes	36

1 INTRODUCTION

En 2017, d'après l'organisation pour l'alimentation et l'agriculture (FAO), le poisson constituait 20% de l'apport moyen en protéine pour 3,3 milliards de personnes. Riche en acide gras polyinsaturés et en micronutriments indispensables à une bonne alimentation, le poisson est reconnu pour ses apports nutritionnels et connaît une popularité croissante auprès des consommateurs. Avec la demande grandissante sur les ressources marines, il est d'autant plus important d'avoir une bonne connaissance de l'écosystème marin, son fonctionnement et son évolution dans le contexte du changement global et des forçages anthropiques. Or, les assemblages microbiens planctoniques jouent un rôle fondamental pour le fonctionnement de l'Océan (Azam *et al.*, 1983). Et leur rôle en tant que producteurs primaires et consommateurs primaires présente une importance capitale pour le reste du réseau trophique, y compris les ressources halieutiques (Pomeroy, 1974). Parmi ces assemblages microbiens, le phytoplancton est un protagoniste majeur. Alors qu'il ne constitue que 0,2% de la biomasse totale des producteurs primaires, il est responsable de près de la moitié de la production primaire globale (Field *et al.*, 1998).

Le changement climatique est susceptible d'altérer rapidement la structure des communautés phytoplanctoniques (Litchman & Klausmeier, 2008). Aujourd'hui plus que jamais, il est crucial de comprendre quels sont les facteurs qui dirigent les assemblages et les dynamiques de ces communautés. De nombreux facteurs sont à considérer, notamment les nutriments, la lumière, l'hydrologie et les interactions biotiques telle que la compétition (Brunet *et al.*, 2006). Pour établir les interactions entre ces différents facteurs et évaluer leur importance relative, de nombreux auteurs suggèrent une approche basée sur les traits des espèces qui forment les communautés (Brunet *et al.*, 2006; McGill *et al.*, 2006; Litchman & Klausmeier, 2008). Parmi ces traits, la taille est une caractéristique non négligeable puisqu'elle influence entre autre, le taux de croissance, l'absorption des nutriments, la flottaison et la réaction au mélange de la colonne d'eau (Brunet *et al.*, 2006). Pour des raisons pratiques, le plancton a traditionnellement été catégorisé par classes de taille, pico-plancton (0.2-2 μ m), nano-plancton (2.0-20 μ m) et micro-plancton (20-200 μ m) (Pan, 2021). Cependant, la taille peut également être utilisée pour définir des groupes fonctionnels (Le Quéré *et al.*, 2005).

L'étude de la distribution des assemblages microbiens requiert le comptage et l'identification de centaines d'espèces dans des milliers d'échantillons. La cytométrie en flux présente l'avantage de permettre l'analyse d'un très grand nombre de cellules en très peu de temps, à la différence des méthodes plus conventionnelles telle que la microscopie (Trask *et al.*, 1982). Dans le cytomètre, les cellules sont entraînées par un liquide vecteur qui les sépare, les aligne et les conduit vers une source lumineuse. Elles

sont interceptées une par une par un ou plusieurs faisceaux laser. Les propriétés optiques de diffusion de la lumière et d'émission de fluorescence des cellules sont alors mesurées. Ces variables permettent de déterminer l'abondance des cellules présentes et de les discriminer en différents groupes. Cependant, ces groupes ne sont pas nécessairement définis par des unités taxonomiques (Trask *et al.*, 1982). Cette méthode permet toutefois de caractériser la structure des communautés planctoniques sur la base de ses groupes fonctionnels (Le Quéré *et al.*, 2005) définis par les classes de taille des cellules.

Pour comprendre comment les écosystèmes marins réagissent à la fois aux contraintes naturelles de l'environnement et aux effets anthropiques, il est nécessaire de conduire des observations à moyen et long terme de l'évolution de l'environnement. L'accumulation des données grâce au développement d'une démarche systématique d'observation des milieux littoraux rend aujourd'hui possible l'étude de séries à long terme. Cette démarche est initialement issue d'initiatives locales, fondées, soit sur des opportunités, soit sur une tradition de culture scientifique propre à chaque station rassemblée aujourd'hui par le Réseau des Stations et Observatoires Marins (RESOMAR). La fédération de ces efforts a notamment conduit à la mise en place du Service d'Observation en Milieu Littoral (SOMLIT), un Service National d'Observation (SNO) qui existe depuis 1995.

Depuis plus de 10 ans, la cytométrie en flux est utilisée par le SNO SOMLIT pour suivre la dynamique temporelle du pico et du nano-phytoplancton à ses différentes stations. À ce jour, 11 stations marines françaises réparties le long du littoral métropolitain font partie du périmètre labellisé du SOMLIT. Parmi elles, trois se situent en Méditerranée : à Banyuls, Marseille et Villefranche. Chaque station acquiert des données selon un protocole national standardisé lors de sorties en mer réalisées tous les 15 jours. Les données acquises lors de mesures directes ou issues des prélèvements d'eau de mer comprennent des variables hydrologiques, des variables biologiques relatives au pico et nano-plancton ainsi que des profils verticaux sur la colonne d'eau (température, salinité, concentration en chlorophylle, turbidité).

La Méditerranée étant particulièrement sensible aux variations naturelles comme à la pression anthropique grandissante (Durrieu de Madron *et al.*, 2011), un intérêt particulier est apporté aux 3 stations méditerranéennes. De plus, celles-ci présentent des conditions environnementales contrastées. De manière générale en Méditerranée, on observe la mise en place d'une thermocline saisonnière du printemps à l'automne. Ce phénomène est amplifié dans le Golfe du Lion où sont situées les stations de Marseille et de Banyuls (Millot, 1990). La circulation de surface dans le Golfe du Lion est dominée par le Courant Nord qui suit le talus continental le long de la côte de Provence et dans le Golfe du Lion jusque la Mer Catalane (Millot, 1990). Les vents dominants, le Mistral dans la région de Marseille et la Tramontane dans la région de Banyuls, sont des vents de nord-ouest. Ils ont des conséquences sur le transport des sédiments ainsi que

sur le déclenchement d'*upwelling* à proximité de la côte. Ces vents peuvent également avoir un impact sur l'étendue du panache du Rhône qui est une source importante d'eau douce en Méditerranée. Cet apport continental peut conduire à la formation d'une couche d'eau douce de surface allant de quelques mètres à une dizaine de mètres de profondeur dans la région sous influence du Rhône (Pairaud *et al.*, 2011). La Méditerranée étant de nature oligotrophe, les apports du Rhône ne sont pas négligeables et peuvent modifier de manière importante la productivité des organismes marins (Pairaud *et al.*, 2011).

Durant ce stage, nous nous intéressons aux assemblages phytoplanctoniques caractérisés par cytométrie en flux depuis 2012 aux trois stations SOMLIT méditerranéennes. Nous cherchons à identifier les ressemblances et les différences entre les stations, aussi bien en matière de structure des communautés qu'en matière de dynamique temporelle. Pour cette étude, les données biologiques relatives au phytoplancton ainsi que les données hydrologiques environnementales associées sont considérées. La particularité de ce stage réside dans l'utilisation de méthodes statistiques multivariées et fonctionnelles. Cette approche présente l'avantage de pouvoir traiter les données temporelles et spatiales apparaissant sous forme de courbes échantillonnées (profils verticaux et séries temporelles). L'aspect fonctionnel permet notamment d'imputer les données manquantes à ce suivi pluriannuel bimensuel.

Dans un premier temps, nous nous intéressons d'une part à la dynamique temporelle des groupes fonctionnels de plancton et d'autre part aux variables indicatrices du réchauffement climatique. Ce, dans l'espoir de mettre en évidence de manière objective les tendances, dans le contexte du changement global. Nous avons ici recours à des méthodes de lissages non-paramétriques et d'analyse de séries temporelles. Dans un second temps, nous cherchons à identifier les différences et les ressemblances entre les trois stations méditerranéennes. Pour cela, nous utilisons des méthodes de classification pour discerner les stations en fonction de la structure de leur communauté et en fonction de leur richesse en nutriments. Enfin, nous souhaitons mettre en évidence les relations et entre les variables biologiques et les variables environnementales. Cette étape est réalisée à l'aide de méthodes d'analyses multivariées.

C'est à notre connaissance la première fois que le jeu de données biologiques de cytométrie SOMLIT sera étudié dans son ensemble, environné par les variables physico-chimiques environnementales associées. Ces analyses, pouvant être réalisées en intégrant les différentes échelles emboîtées (intra et inter annuelles) (Nerini *et al.*, 2021) permettrait de comprendre non seulement la dynamique des communautés de phytoplancton dirigée par les conditions environnementales au sein d'un site, mais aussi d'appréhender l'impact du changement global sur la structure de ces communautés.

2 MATÉRIEL & MÉTHODES

2.1 SOMLIT (Service d'Observation en Milieu LITtoral)

2.1.1 Les stations

La station de Banyuls, Sola, se situe en sortie du Golfe du Lion. La station est située sur un fond sablo-vaseux sur 27m de profondeur à 0,5 mille de la côte. Comme le suggère la localisation des fleuves à proximité de Banyuls (Figure 1), cette station est à la fois sous influence de la circulation générale et des apports continentaux.

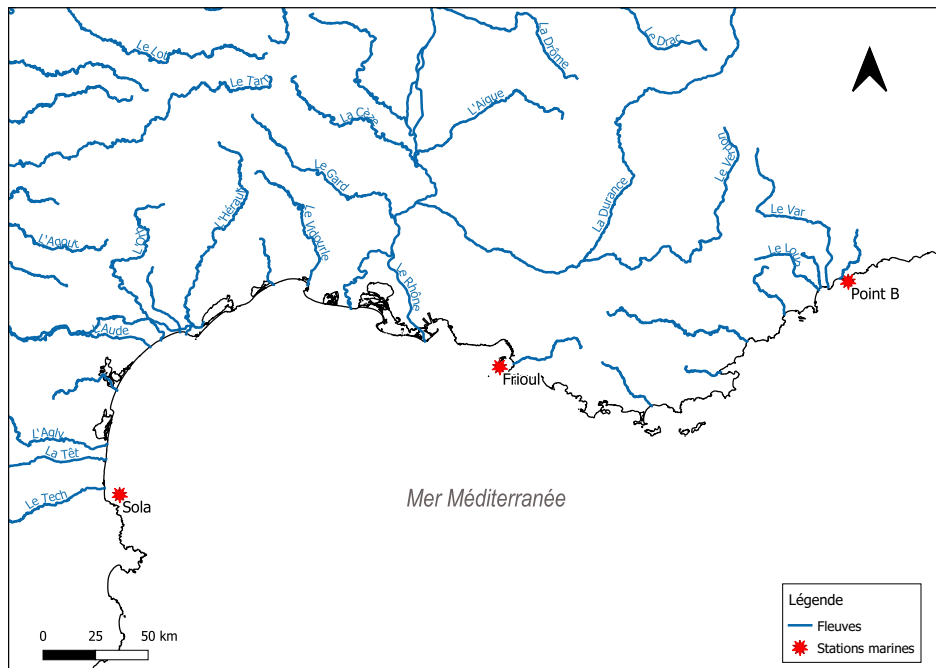


Figure 1 : Carte des stations SOMLIT méditerranéennes : Sola à Banyuls, Frioul à Marseille et Point B à Villefranche.

La station de Marseille, Frioul, se situe sur un fond sableux, sur 60m de profondeur à 5km de la côte. Cette station localisée dans le golfe de Marseille est considérée comme strictement marine. Cependant, la proximité du Rhône, l'un des fleuves les plus importants de la Méditerranée, fait que cette station subit une influence continentale marquée.

La station de Villefranche, Point B, est située dans la Baie du Villefranche sur 80m de fond. Elle est abritée des vents d'est et d'ouest dominants dans la région. Les masses d'eau du large pénètrent dans la rade grâce au Courant Nord et à l'absence de plateau continental. Les apports d'eau douce les plus proches sont les fleuves Roya, Paillon et Var, avec de fortes variations saisonnières. Le site est également soumis à des influences anthropiques d'amplitude saisonnière. Il subit notamment un trafic maritime intense.

2.1.2 Prélèvements et protocoles

L'acquisition des données est réalisée par des membres du SOMLIT lors de sorties en mer tous les 15 jours à Marseille et de manière hebdomadaire à Villefranche et à Banyuls. Depuis la création du SNO, des protocoles nationaux ont été établis progressivement. Aujourd'hui, les protocoles de prélèvements et d'analyses sont standardisés pour toutes les stations du SOMLIT. Cela permet une comparaison des observations entre les stations facilité et plus robuste.

À chaque sortie en mer, l'eau est prélevée dans des bouteilles Niskin en surface et au fond à des profondeurs définies pour chaque station. Elles sont de 3m en surface et 24m de profondeur à Sola, 1m et 55m au Frioul, 1m et 50m au Point B. La température et la salinité sont alors mesurées. L'oxygène dissous est déterminé par la méthode de Winkler. De l'eau est échantillonnée des bouteilles Niskin pour réaliser une mesure en laboratoire du pH par spectrophotométrie. Un échantillonnage est également réalisé pour l'analyse de l'ammonium (NH_4) par fluorimétrie. D'autres échantillonnages des bouteilles Niskin sont destinés à l'analyse des nutriments, nitrates (NO_3), nitrites (NO_2), phosphates (PO_4) et silicates ($\text{Si}(\text{OH})_4$) par la méthode de Aminot et Kerouel (2004). La concentration de la matière en suspension (MES) est déterminée après filtration d'un échantillon dans des filtres de $0,7\mu\text{m}$. Les concentrations en carbone organique particulaire (COP) et en azote organique particulaire (NOP) sont déterminées par combustion du matériel particulaire après filtration d'un échantillon. La chlorophylle *a* est elle, déterminée par fluorimétrie d'un échantillon. Les protocoles détaillés sont disponibles sur le site du SOMLIT (<https://www.somlit.fr/>).

Les profils verticaux sont réalisés avec une sonde *conductivity-temperature-depth* (CTD SBE19plusV2, Seabird). Elle permet de mesurer la température et la salinité dérivée de la conductivité. Couplée à un capteur de lumière biosphérique PAR, la radiation lumineuse est également mesurée par la CTD. Associée à un fluorimètre, elle permet aussi d'obtenir le profil de fluorescence en relation avec la concentration en chlorophylle *a*. La sonde CTD est déployée entre la surface et le fond, en station, à chaque sortie en mer. Dans le cas de Villefranche, la sonde est déployée au Point B+ sur 150m de profondeur.

L'eau prélevée à la surface dans une bouteille Niskin est également échantillonnée pour les analyses du pico et nano-plancton (Marie *et al.*, 2014). L'échantillon est fixé avec un mélange de Glutaraldéhyde et Polaxamer, et est conservé jusqu'à son envoi à la plate-forme de Banyuls où tous les échantillons sont analysés par cytométrie en flux (Marie *et al.*, 2014).

Lors de l'analyse par cytométrie, les cellules en suspension dans l'eau de mer filtrée passent une à une au travers d'un ou de plusieurs faisceaux laser. Les signaux de diffusion et de fluorescence émis par

chaque cellule sont collectés par des détecteurs. La fluorescence émise par les cellules photosynthétiques provient principalement de leurs pigments photosynthétiques. Les longueurs d'ondes émises dépendent du type et de la concentration des pigments. La présence de chlorophylle induit une émission de fluorescence dans le rouge et la présence de phycoérythrine induit une émission dans l'orange. Dans le cas de cellules hétérotrophes, la seule fluorescence émise est la fluorescence verte induite par le marquage spécifique de l'ADN des cellules par le colorant SYBR Green I (Molecular probes). La diffusion lumineuse à 90 degrés est indicatrice de la structure (granularité) de la cellule (Trask *et al.*, 1982) et la diffusion de la lumière aux petits angles est en relation avec la taille de la particule (Dubelaar *et al.*, 2007).

Les propriétés optiques des cellules mesurées par cytométrie en flux permettent de distinguer des groupes, formés par des cellules aux propriétés homogènes, dans la communauté pico et nano-planctonique. Néanmoins, ces groupes ne contiennent pas nécessairement chacun une unité taxonomique définie. Il est donc possible de distinguer les bactéries hétérotrophes, et parmi elles, celles à haut contenu en acide nucléique (bactéries HNA) et celles à faible contenu en acide nucléique (bactéries LNA). Parmi les organismes autotrophes il est possible de distinguer les *Prochlorococcus* Chisholm *et al.*, 1992, les *Synechococcus*, les Cryptophytes Pascher, 1913, les pico-eucaryotes et les nano-eucaryotes (Marie *et al.*, 1999).

Les *Prochlorococcus* et les *Synechococcus* sont des procaryotes autotrophes. Ce sont des producteurs primaires particulièrement importants dans des conditions oligotrophes. Ils constituent des proies pour les Ciliés Copeland, 1956, les Dinoflagellés Bütschli, 1885 et le nano-plancton hétérotrophe (Romagnan *et al.*, 2015). Les pico-eucaryotes autotrophes, d'une taille inférieure à $2\mu\text{m}$, présentent une forte affinité pour les nutriments et la lumière, qui grâce à un ratio surface/volume élevé leur confère un avantage en conditions oligotrophes (Le Quéré *et al.*, 2005). Les Cryptophytes sont des eucaryotes de taille hétérogène. Les nano-eucaryotes englobent tous les autres organismes autotrophes avec une taille comprise entre 2 et $20\mu\text{m}$ (Figure 2).

2.1.3 Jeux de données

Les données collectées lors des campagnes du SOMLIT sont hébergées dans la base de données SOMLIT (<https://www.somlit.fr/>). Les données hydrologiques, biologiques (cytométrie) et CTD pour les trois sites d'intérêt, Banyuls, Marseille et Villefranche, ont été téléchargées depuis le site internet. Le Tableau 1 résume les données récupérées pour les trois stations. Le jeu de données hydrologiques (HYDRO) contient les données de température, salinité, oxygène, pH, NH_4 , NO_3 , NO_2 , PO_4 , Si(OH)_4 , COP, NOP, MES et chlorophylle *a*. Le jeu de données biologiques (PICONANO) contient les données de cytométrie

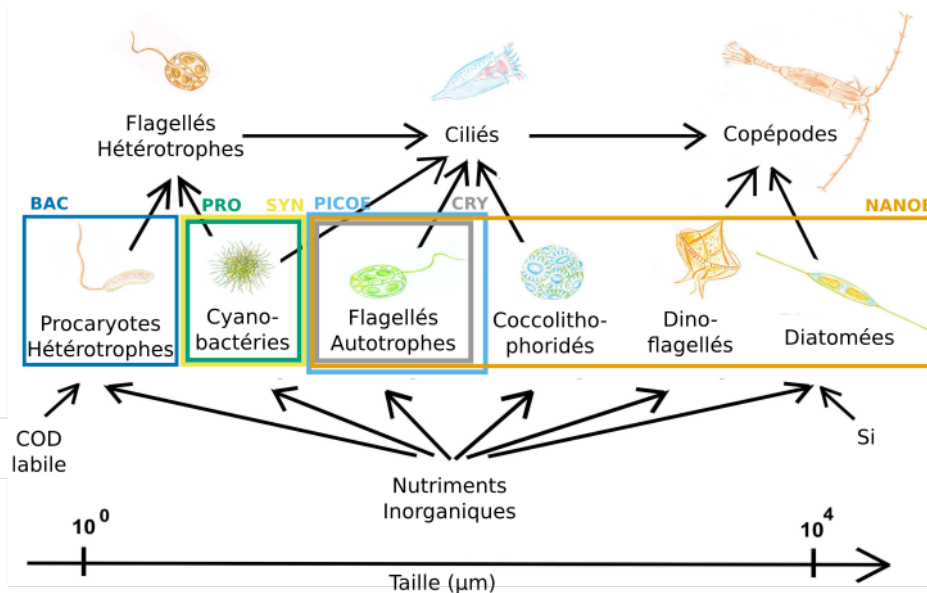


Figure 2 : Réseau trophique planctonique pélagique. Les groupes distingués par cytométrie en flux sont représentés par des cases qui englobent les organismes qu'ils peuvent contenir, en bleu foncé les bactéries hétérotrophes (BAC), en vert les *Prochlorococcus* (PRO), en jaune les *Synechococcus* (SYN) en bleu clair les pico-eucaryotes (PICOE), en gris les Cryptophytes (CRY) et en orange les nano-eucaryotes (NANOE). Modifié depuis Våge & Thingstad (2015).

des groupes de plancton présentés précédemment. Pour tous les groupes, les données d'abondance et de diffusion lumineuse moyenne des cellules sont disponibles. Pour le phytoplancton, sont de même disponibles les intensités moyennes d'auto-fluorescence rouge, et pour les bactéries hétérotrophes les intensités moyennes de fluorescence verte. Les valeurs d'auto-fluorescence orange de *Synechococcus* et des Cryptophytes sont également renseignées. Le jeu de données CTD contient les profils verticaux de la surface au fond pour les variables température, salinité, chlorophylle *a* et éclaircissement.

Une première visualisation des données permet d'identifier les données aberrantes. La Figure 3 présente les données de surface des variables hydrologiques au cours du temps toutes stations confondues. Les éventuelles valeurs aberrantes sont identifiées en rouge. La ligne pointillée orange correspond au début du jeu de données cytométriques PICONANO et la lignée pointillée bleue correspond à la mise en place du protocole national pour la variable présentée. Étant donné la mise en place tardive d'un protocole standardisé pour le pH, cette variable n'est pas analysée. La mesure de l'oxygène étant sujette à de nombreux biais liés à la précision de l'échantillonnage, celle-ci est également ignorée. La Figure 4 présente les données de température de surface et au fond pour les trois stations. Les graphiques nous permettent d'identifier une potentielle inversion des données de fond et de surface à Villefranche entre le début de la série temporelle et 2002. Comme il est impossible de savoir avec certitude quelles observations ont été interverties, les données HYDRO de Villefranche datant d'avant 2002 ne sont pas exploitées.

Tableau 1 : Données téléchargées à partir du site internet du SOMLIT pour les trois stations méditerranéennes. Résume le nombre d'observations disponibles (N) pour chaque site en fonction de la profondeur de l'échantillonnage (Prof).

Station	Données HYDRO			Données PICONANO			Données CTD		
	Début	N	Prof (m)	Début	N	Prof (m)	Début	N	Prof (m)
Banyuls	1997	1183	3	2012	215	3	2019	93	3-24
	1997	1184	24						
Marseille	1997	583	1	2012	213	1	1994	627	0-55
	2004	407	55						
Villefranche	1995	1320	1	2012	192	1	2007	634	1-150
	1995	1321	50						

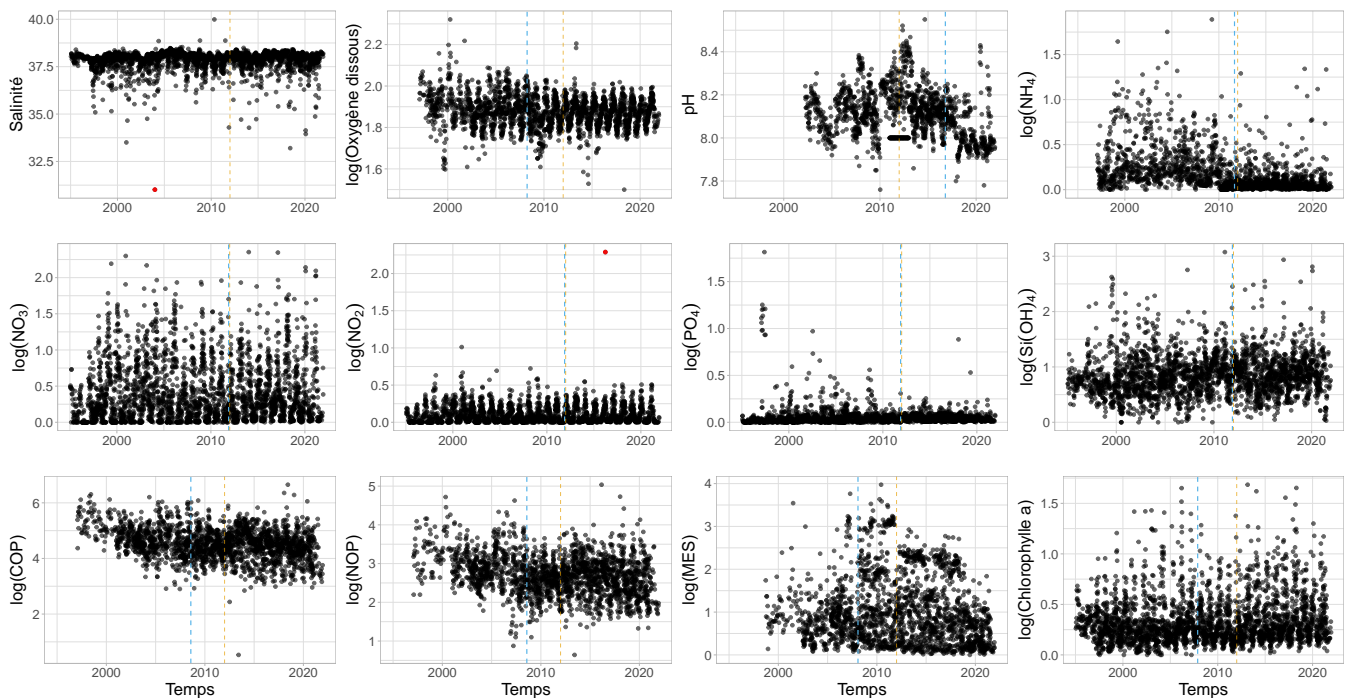


Figure 3 : Séries temporelles des données hydrologiques de surface pour les trois stations SOMLIT méditerranéennes confondues. La ligne pointillé bleue correspond à la mise en place du protocole national pour la variable présentée. La ligne pointillée orange correspond au début du jeu de données PICONANO. Les éventuelles données aberrantes sont représentées en rouge.

Dans le cadre de ce stage, nous nous sommes focalisés sur le phytoplancton. Par soucis de temps, les bactéries hétérotrophes ne sont pas prises en compte. La combinaison des propriétés optiques des cellules mises en évidence par la cytométrie en flux, permet de distinguer des groupes dans la communauté phytoplanctonique. La Figure 5 représente avec une échelle logarithmique, la valeur moyenne d'auto-

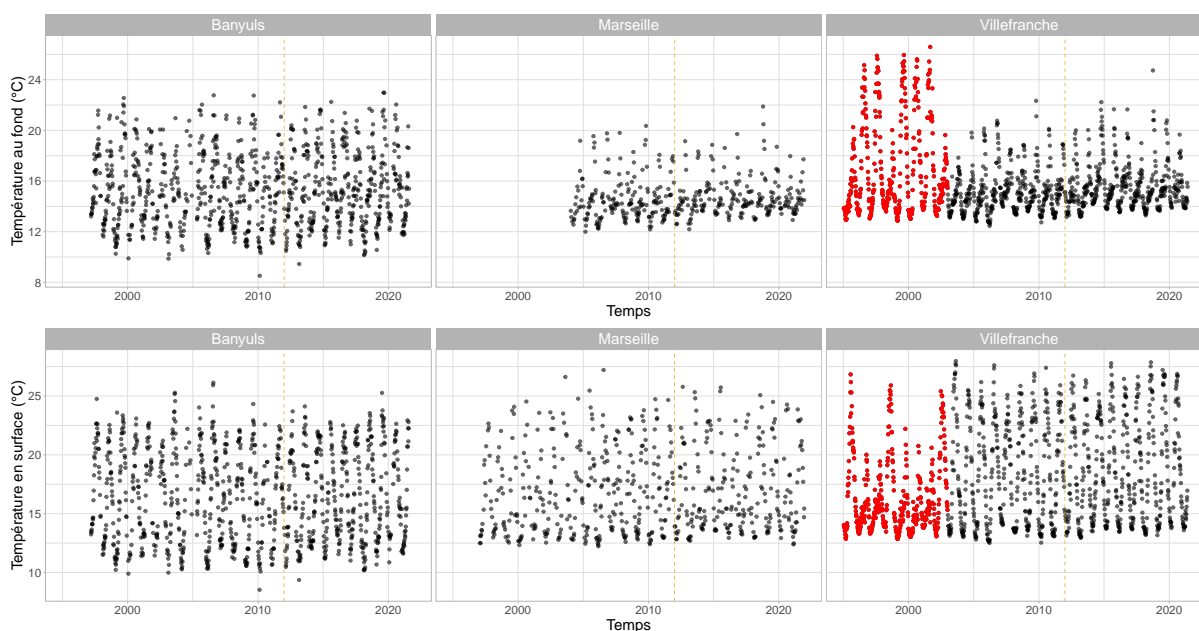


Figure 4 : Séries temporelles des données de température en surface et au fond pour les trois stations méditerranéennes. La ligne pointillée orange correspond au début du jeu de données PICONANO. Les éventuelles données aberrantes sont représentées en rouge.

fluorescence rouge d'un groupe lors d'un prélèvement, en fonction de la valeur moyenne de diffusion lumineuse. À l'exception d'une observation pour le groupe de *Synechococcus* isolée de toutes les autres, les données paraissent fiables puisque l'on identifie facilement les différents groupes.

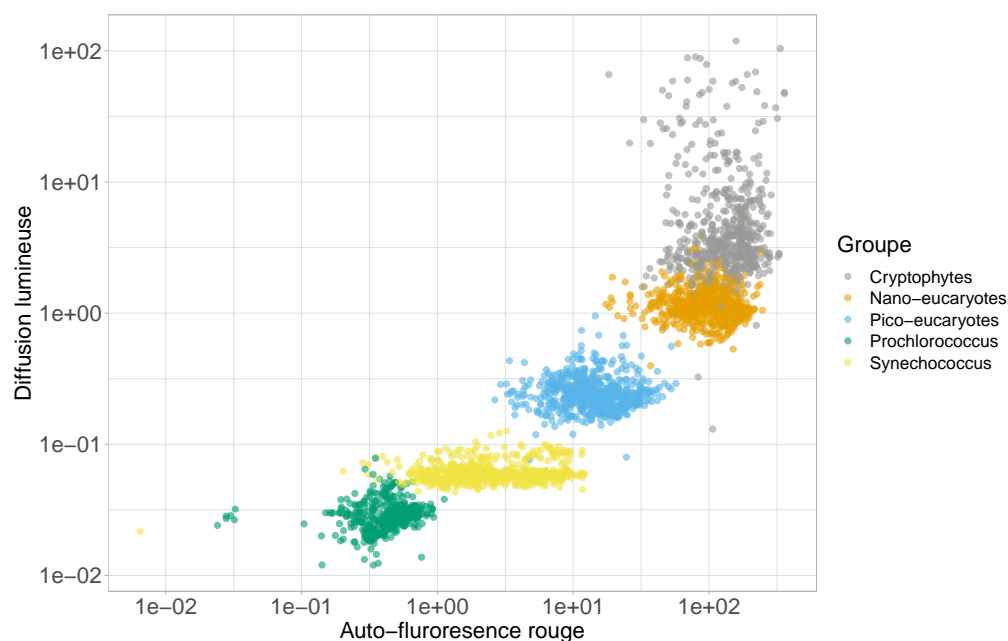


Figure 5 : Auto-fluorescence rouge moyenne *versus* diffusion lumineuse à 90° moyenne de chaque groupe de phytoplancton pour chaque prélèvement pour les trois stations SOMLIT méditerranéennes.

En ce qui concerne les données PICONANO, durant ce stage l'étude s'est focalisée sur les variables d'abondance et de diffusion lumineuse (SSC), cette dernière étant indicatrice de la structure des organismes.

2.2 Analyse de séries temporelles

La visualisation et l'étude de l'évolution des populations au cours du temps sur toute la durée de la période d'étude peut constituer une première étape pour déceler les tendances et comprendre la dynamique temporelle que suit chaque population. De plus, dans le contexte du réchauffement, il convient d'étudier la tendance que suit la température au cours du temps au sein des stations. La durée d'observation des données de 10 ans pour les données biologiques et plus pour les données hydrologiques, rend possible l'étude de l'évolution des variables au cours du temps. Ces variables peuvent donc être étudiées avec des méthodes d'analyse de séries temporelles.

2.2.1 Lissage non-paramétrique

Afin d'explorer les séries temporelles pour identifier une structure et déceler une éventuelle tendance, dans un premier temps, un lissage a été réalisé sur chaque variable d'intérêt et pour chaque station. Le choix du lissage non paramétrique est fait car nous ne connaissons pas la structure des séries et que c'est justement ce que nous cherchons à identifier.

Un modèle de régression non-paramétrique peut s'écrire de la manière suivante :

$$y_i = m(x_i) + \varepsilon_i$$

Avec $m(x)$ la courbe de régression qui correspond à l'espérance conditionnelle $E(Y|X = x)$. Il est possible de définir un estimateur à noyau de cette courbe. De telle manière qu'aux valeurs de y_i correspondant aux points x_i les plus proches de x est accordé un poids plus important. Dans le cas où le voisinage est large, il convient d'utiliser des polynômes d'ordre p . Dans ce cas, un estimateur par régression polynomiale locale $\hat{m}_p(x)$ correspond à $\hat{\beta}_0$ qui minimise l'équation suivante (Simonoff, 1996).

$$\sum_{i=1}^n (y_i - \beta_0 - \dots - \beta_p(x - x_i)^p)^2 K\left(\frac{x - x_i}{h}\right)$$

On fait le choix d'utiliser un noyau $K()$ gaussien, non-borné qui permet d'avoir une variance conditionnelle finie (Simonoff, 1996). L'ordre des polynômes est fixé à 2 car un degré supérieur n'apporte pas

une précision significativement meilleure. Le choix de la fenêtre de lissage h est fait en testant plusieurs valeurs. Nous conservons celle qui permet une bonne estimation des observations réelles tout en produisant une régression suffisamment lisse.

L'intervalle de confiance à 95% de la régression est calculé à partir de la méthode décrite par Fan & Gibels (1996). Cette méthode se base sur l'expression asymptotique du biais et de la variance conditionnels. Dans cette étude, ils sont approximés par un développement de Taylor d'ordre 1.

2.2.2 Saisonnalité

Pour déterminer la saisonnalité dans les séries temporelles, nous nous sommes intéressés à la densité spectrale. La densité spectrale est analogue à une densité de probabilité. Elle donne la même information que la fonction d'auto-covariance. Cependant, au lieu d'exprimer la covariance en terme de *lag*, elle l'exprime en terme de cycle. La densité spectrale peut être estimée par un périodogramme.

Le périodogramme échelonné $P(\omega)$ indique quelles composantes de fréquences ont une grande magnitude et lesquelles ont une faible magnitude. De grandes valeurs de $P(\omega)$ indiquent donc quelles fréquences ω sont prédominantes dans la série. Les petites valeurs de $P(\omega)$ sont plus susceptibles d'être associées au bruit.

La première étape consiste à s'affranchir de la tendance présente dans la série. En effet, les tendances vont donner de l'importance à des composantes de basses fréquences qui vont obscurcir les plus hautes fréquences. Nous réalisons donc un centrage des données de la forme $x_t - \bar{x}$ en amont de l'analyse. Le périodogramme peut ensuite être obtenu avec la transformation de Fourier rapide (FFT) (Shumway & Stoffer, 2016).

Cependant, le périodogramme est sujet à de nombreuses incertitudes. Pour que le périodogramme soit un estimateur fiable de la densité spectrale, il convient de lisser le périodogramme (Shumway & Stoffer, 2016).

Le périodogramme lissé, ou périodogramme moyen $\bar{f}(\omega)$, est défini comme la moyenne des valeurs du périodogramme sur la bande de fréquence \mathcal{B} . Les valeurs du spectre à l'intérieur de cette bande doivent être relativement constantes pour que le spectre lissé soit un bon estimateur. La largeur de cette bande est ajustée par le choix de la fenêtre. Nous pouvons obtenir une approximation d'un intervalle de confiance à $100(1 - \alpha)\%$ de la valeur du spectre $f(\omega)$ de la forme (Shumway & Stoffer, 2016) :

$$\frac{df\bar{f}(\omega)}{\chi_{df}^2(1 - \alpha/2)} \leq f(\omega) \leq \frac{df\bar{f}(\omega)}{\chi_{df}^2(\alpha/2)}$$

Il est commun d'observer des pics mineurs au niveau des harmoniques car la plupart des signaux ne sont pas des sinusoides parfaits. Les harmoniques permettent alors de capturer le comportement non-sinusoidal du signal. La question est de décider si un pic est significatif ou non. En général, un pic est considéré significatif si sa borne inférieure de l'intervalle de confiance est supérieure au bruit.

Lors de la création de la fenêtre, il est possible que certaines fréquences en dehors de l'intervalle voulu contaminent la bande. Pour supprimer cet effet de fuite, il est possible d'utiliser des *taper*. Les *taper* ont généralement une forme qui amplifie le centre des données par rapport aux extrémités, comme par exemple une cloche cosinus (Shumway & Stoffer, 2016).

Le périodogramme moyen est calculé à l'aide du paquet *astsa*. Il permet d'ajouter un taper de la forme d'une cloche cosinus et d'ajuster le noyau. Le noyau choisi est celui qui produit un périodogramme moyen le plus ressemblant au périodogramme échelonné.

2.3 Analyse en Composantes Principales Fonctionnelle

Une Analyse en Composantes Principales (ACP) classique a pour objectif de trouver les combinaisons linéaires de variables quantitatives qui maximisent la variance au sein d'un jeu de données. Notre intérêt se porte, non pas sur des observations ponctuelles, mais sur l'évolution dans le temps ou dans l'espace de ces observations. Cette évolution peut être modélisée par des fonctions. Si une fonction peut être exprimée par des coefficients, il suffit de chercher alors les combinaisons linéaires de ces coefficients qui maximisent la variance entre ces fonctions. C'est le principe de l'ACP fonctionnelle.

2.3.1 Lissage paramétrique

La première étape consiste à transformer des données d'échantillonnage discrètes en des fonctions continues sur l'intervalle que nous souhaitons comparer. Dans le cas des données HYDRO et PICAONANO nous nous intéressons à des unités annuelles, une fonction correspond donc à une année. Dans le cas des données CTD, une fonction correspond à un profil vertical de mesures à une date d'échantillonnage.

Pour obtenir une fonction $x(t)$, un système de fonctions de bases est utilisé. C'est un ensemble de K fonctions ϕ_k connues et indépendantes les unes des autres. Une association linéaire de ces K fonctions permet d'approximer notre fonction. La fonction $x(t)$ s'exprime alors comme une expansion linéaire :

$$x(t) = \sum_{k=1}^K c_k \phi_k(t)$$

Les coefficients d'expansion sont notés c_k .

La méthode de lissage choisie est celle par pénalisation de la courbure ou régularisation. Cette méthode de lissage permet de choisir le paramètre de lissage. Le paramètre de lissage λ nous permet de contrôler l'équilibre entre l'ajustement de la courbe lissée aux données initiales et la variabilité de cette courbe. Une méthode subjective est utilisée : elle consiste à tester plusieurs valeurs de λ et à sélectionner celle pour laquelle le lissage est le plus adapté. Des splines cubiques ont été choisies comme fonctions de bases, comme suggéré par Ramsay & Silvermann (2005).

2.3.2 ACP fonctionnelle

L'ACP fonctionnelle est réalisée avec le paquet *fda*. À la différence d'une ACP classique, au lieu de disposer de réalisations de J variables, pour N échantillons, nous allons étudier N fonctions $x_i(t)$. Ce sont les coefficients d'expansion des fonctions $x_i(t)$ qui font office de variables. Dans le cas où il manque une moyenne mensuelle pour plusieurs années, l'ACP est réalisée avec le paquet *fdapace* qui supporte les valeurs non attribuées.

2.3.3 Perturbation de la moyenne

Pour interpréter visuellement les résultats de l'ACP fonctionnelle, une solution est d'exprimer ses composantes principales comme perturbation de la moyenne $\hat{\mu}(t)$. Soit λ_l la valeur propre de la l -ième composantes principales et $\hat{\xi}_l(t)$ sa fonction propre associée, il est possible de les exprimer de la manière suivante (Nerini *et al.*, 2021) :

$$\hat{\mu}(t) \pm \sqrt{\lambda_l} \times \hat{\xi}_l(t)$$

2.4 Analyse Factorielle Discriminante

L'Analyse Factorielle Discriminante (AFD) est une méthode de classification supervisée. C'est à dire qu'elle permet de classer des observations en K classes connues. Elle peut être utilisée comme une méthode descriptive qui permet d'identifier les variables quantitatives qui discriminent au mieux les K modalités d'une variable qualitative. Cette méthode est utilisée pour tenter de discerner les trois stations méditerranéennes ($K = 3$) en fonction de l'abondance de chaque groupe de phytoplancton, de la taille des organismes de chaque groupe de phytoplancton ou de la concentration de chaque nutriment, au cours d'une année. Le nombre d'axes discriminants d'une AFD est inférieur ou égale à $K - 1$. Il est alors possible de visualiser les résultats de l'AFD en deux dimensions, en ayant toutes les informations.

Une étape préliminaire à l'AFD consiste à passer des observations ponctuelles au cours d'une année à une observation qui correspond à l'évolution d'une variable au cours d'une année. Cette transformation est réalisée à l'aide d'ACP fonctionnelles présentées précédemment. Une ACP est réalisée pour chaque variable d'intérêt. Les variables qualitatives utilisées pour l'AFD correspondent aux scores des composantes principales des ACP. Selon le pourcentage d'inertie des composantes principales, une ou deux composantes principales sont récupérées pour chaque ACP. Si deux composantes principales sont récupérées, la deuxième est pondérée en fonction de son pourcentage d'inertie.

Nous obtenons ainsi de nouvelles observations correspondant chacune à une année pour un site. Pour chaque observation, il y a $p \times j$ variables quantitatives qui correspondent aux scores des p composantes principales de j ACP.

Dans le cas où l'on a des valeurs non-attribuées, la fonction *imputePCA* du paquet *missMDA* est utilisée. Cette fonction permet de simuler les valeurs manquantes en ajustant une valeur initiale, qui correspond à la moyenne de la variable, à l'aide d'ACP successives. Cette méthode peut être utilisée comme méthode préliminaire à une ACP (Josse & Husson, 2012). Or, une AFD est assimilable à une ACP particulière, avec ici 2 axes discriminants ($K - 1$). Le nombre de composantes principales utilisées pour prédire les valeurs manquantes est donc fixé à 2.

Une fois le tableau d'observation centré et réduit, l'AFD est réalisée avec la fonction *lda* du paquet *MASS*.

2.5 Analyse Canonique des Corrélations

L'Analyse Canonique des Corrélations (ACC) est une méthode qui permet d'explorer les corrélations entre deux jeux de données quantitatives issues du même échantillonnage (González *et al.*, 2008). Les variables canoniques sont des combinaisons linéaires de variables d'un jeu de données. Une paire de variables canoniques est constituée d'une variable canonique venant du premier jeu de données et d'une autre venant du second jeu de données. L'ACC fonctionne comme un algorithme itératif visant à trouver les combinaisons linéaires qui maximisent la corrélation entre ces variables canoniques (corrélation canonique) (González *et al.*, 2008).

L'ACC est utilisée pour explorer les corrélations entre les données biologiques et les données hydrologiques. Dans un premier temps, nous nous intéressons aux relations entre les données d'abondance et les concentrations en nutriments, dans un second temps à un possible lien entre les valeurs de SSC et les concentrations en nutriments.

Comme pour l'AFD, nous allons nous intéresser à l'évolution d'une variable au cours d'une année. Nous avons donc recours aux scores des ACP utilisés lors de l'AFD. Au cours des analyses, nous avons remarqué que, souvent, la première composante principale rend compte d'un effet moyen. Si c'est en effet le cas, nous récupérons uniquement le score de la première composante principale. Le premier jeu de données $n \times p$ contient n observations correspondant à une année, à un site. Et p variables quantitatives qui correspondent aux scores des premières composantes principales issues des ACP pour chaque groupe de phytoplancton. Dans le premier cas il s'agit des variables d'abondance et dans le second cas des variables de diffusion lumineuse. Le deuxième jeu de données $n \times q$ contient n observations correspondant aux mêmes années aux mêmes sites que le premier jeu de données. Les q variables quantitatives correspondent cette fois aux scores des premières composantes principales issues des ACP pour chaque nutriment.

L'ACC est réalisée avec le paquet *CCA* qui supporte les valeurs manquantes.

3 RÉSULTATS

3.1 Variations dans le temps

3.1.1 Abondance des groupes phytoplanctoniques

Comme nous disposons de données échantillonnées plus ou moins régulièrement deux fois par mois, nous souhaitons lisser les données pour avoir une évolution continue de l'abondance des groupes de phytoplancton au cours du temps.

Pour chaque groupe de phytoplancton défini par cytométrie, et pour chaque station, une régression polynomiale locale sur les observations d'abondance est réalisée pour transformer les données discrètes en fonctions continues. Après avoir essayé plusieurs fenêtres de lissage, celle qui convient le mieux pour chacune des régressions est une fenêtre de 40 jours. Les lissages obtenus sont présentés en Figure 6. Toutes les séries présentent une saisonnalité annuelle plus ou moins marquée.

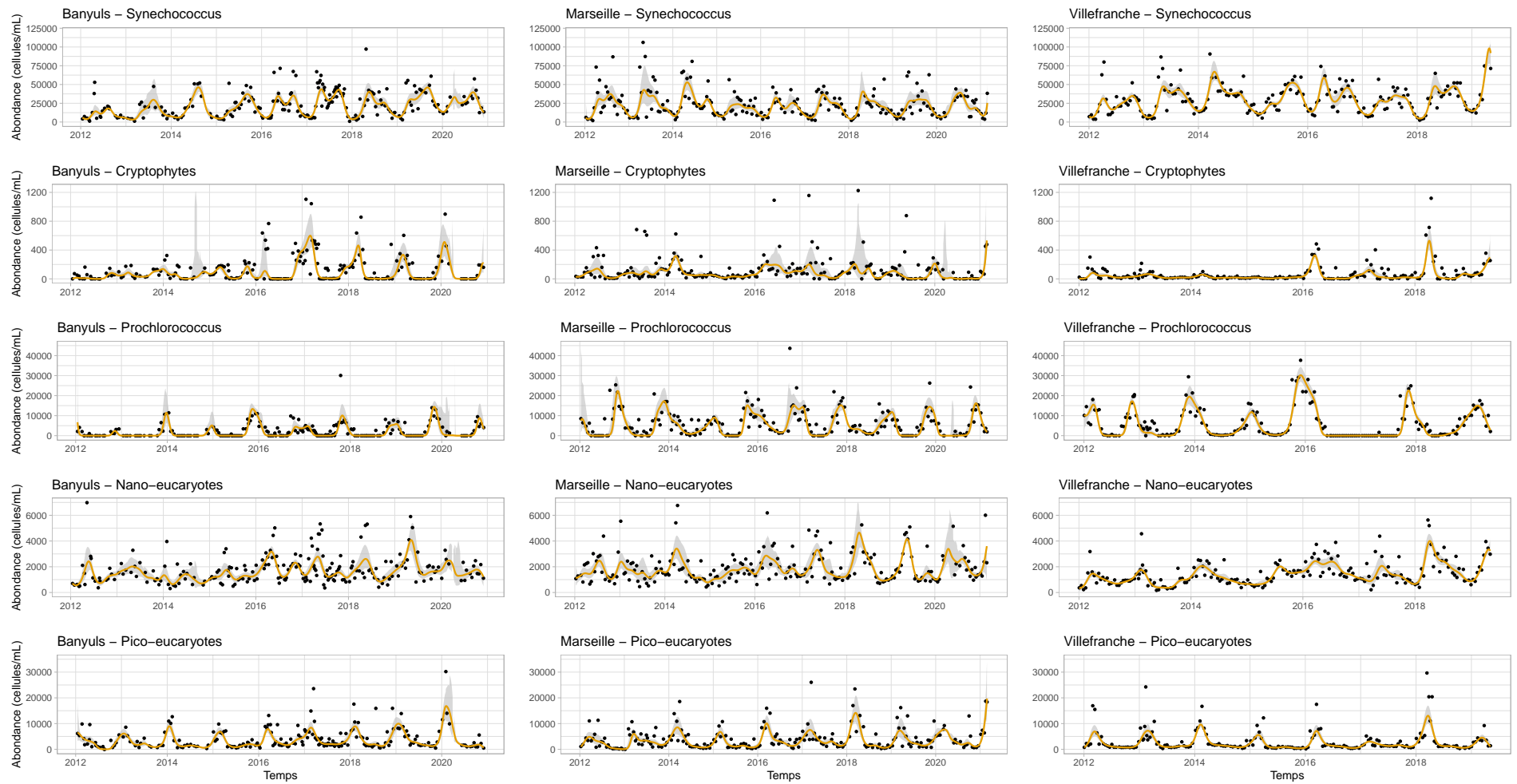


Figure 6 : Séries temporelles des abondances des groupes de phytoplancton pour les trois stations méditerranéennes. Les points noirs correspondent aux données brutes, la courbe orange à la régression quadratique locale avec noyau gaussien avec une fenêtre de 40 jours. Elle est associée à son intervalle de confiance à 95% en gris.

Pour caractériser la saisonnalité, le périodogramme de chaque régression est réalisé. Le périodogramme moyen des *Synechococcus* à Banyuls est présenté en Figure 7. Celui-ci identifie clairement deux maxima à des fréquences de 1.02Δ et 2Δ . Le pic à la fréquence 1.02Δ correspond environ à 1 cycle par an, soit une saisonnalité annuelle. Le pic à la fréquence 2Δ correspond à 2 cycles par an, soit une saisonnalité de 6 mois. Les bornes inférieures de l'intervalle de confiance à 95% des deux pics sont supérieures au reste du signal. Ils peuvent donc être considérés comme significatifs. En revanche, la petite bosse qui ressort à une fréquence de 2.97Δ n'est pas significative. Pour les autres séries, à l'exception des nano-eucaryotes à Villefranche, les périodogrammes permettent de mettre en évidence une saisonnalité annuelle significative. Les nano-eucaryotes à Villefranche présentent des cycles de 2 ans. Pour les pico-eucaryotes et les *Synechococcus* aux trois stations, et pour les nano-eucaryotes à Banyuls, une saisonnalité significative de 6 mois est aussi mise en évidence.

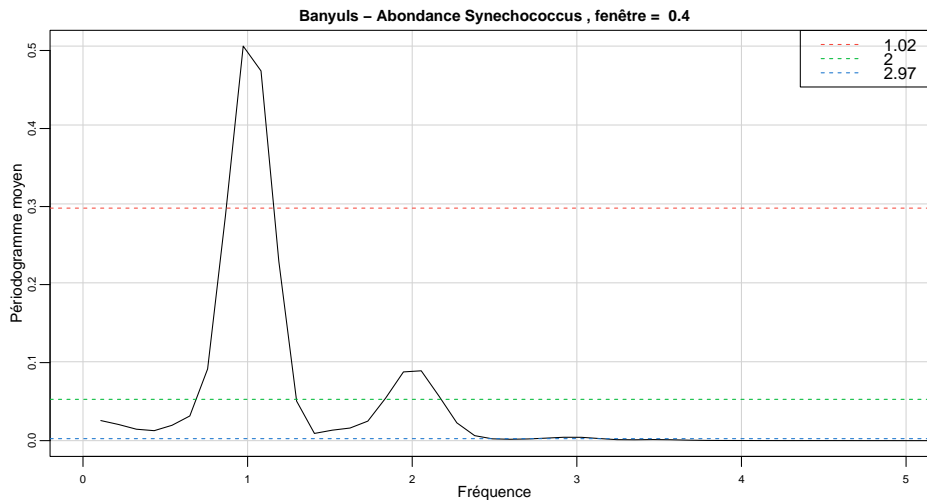


Figure 7 : Périodogramme moyen de la série d'abondance de *Synechococcus* à Banyuls. L'axe de fréquences est gradué par multiple de $\Delta = 1\text{cycle}/1\text{an}$. Les lignes pointillées correspondent à la borne inférieure de l'intervalle de confiance à 95% des pics identifiés dans la légende.

3.1.2 Température

Dans le contexte du réchauffement climatique, nous cherchons d'abord à voir comment il se fait ressentir au sein des stations. Une des analyses qui ont été réalisées se focalise sur l'évolution au cours du temps de la température de la colonne d'eau à Marseille.

Les données CTD des profils de température à Marseille ne sont pas observées à intervalle de profondeur régulier. Il y a environ une observation tous les 0,2 ou 0,3m. Une interpolation est réalisée pour avoir des données régulières tous les 0,5m de 0,5m à 55m de profondeur. Ce qui représente 112 points

pour chaque profils verticaux. Le choix est fait de prendre deux fois moins de fonctions de bases que de points. Pour obtenir un profil lisse qui ne présente pas trop d'ondulations, un lissage avec pénalisation de la courbure avec 55 fonctions de bases est réalisé. Plusieurs paramètres de lissage λ sont testés et le lissage obtenu avec chaque paramètre est visualisé pour 8 profils choisis au hasard. Nous faisons le choix de conserver $\lambda = 0,1$ qui produit le lissage visualisé en Figure 8.

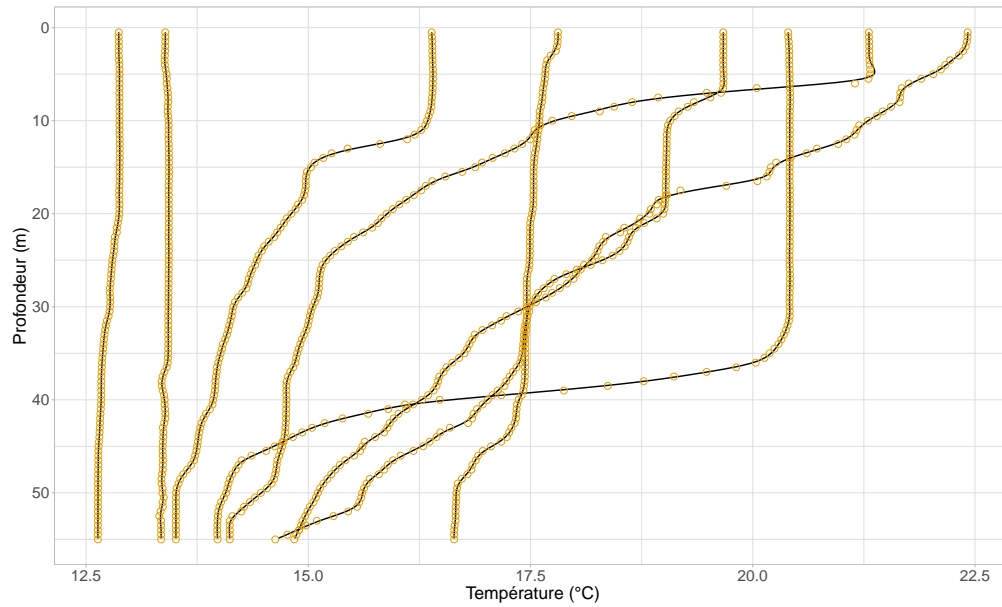


Figure 8 : Profils verticaux de température à Marseille après lissage avec pénalisation, $\lambda = 0,1$, à partir des points oranges pour 8 dates choisies au hasard.

À partir des courbes lissées, l'ACP fonctionnelle est réalisée sur tous les profils de température à Marseille. Les deux premières composantes principales de cette ACP résument à elles deux 97,78% des variations observées entre toutes les courbes. L'effet de ces deux premières composantes est présenté sur la Figure 9a). Un écart positif à la moyenne, représentée par la courbe noire, signifie que le profil se rapproche de la courbe représentée par des croix grises. Un écart négatif à la moyenne signifie que le profil se rapproche de la courbe représentée par des tirets oranges. Sur une visualisation bi-dimensionnelle des scores de l'ACP, cette moyenne correspondrait à l'origine des axes. La première composante principale (PC1), qui représente 89,7% des variations, représente un effet moyen. Une valeur positive sur l'axe 1 indique que la température moyenne du profil est supérieure à la moyenne de tous les profils. À l'inverse, une valeur négative sur l'axe 1 implique que la température moyenne du profil lui est inférieure. La deuxième composante principale (PC2), qui représente 8,08% des variations, représente un effet de stratification. Une valeur positive sur l'axe 2 implique que la température est plutôt homogène le long de la colonne d'eau. Une valeur négative sur l'axe 2 suggère une forte diminution de la température à mesure

que la profondeur augmente. La Figure 9 b) représente sous forme de séries temporelles les scores des deux premières composantes principales pour chaque profil CTD observé. Une saisonnalité annuelle est observable pour les deux composantes. La PC1 met en évidence une température moyenne plus basse la première partie de l'année. La PC2 quant à elle met en évidence une forte stratification de la colonne d'eau en milieu d'année. À partir de 2003 cette stratification semble plus importante puisque des valeurs négatives plus extrêmes sont observées.

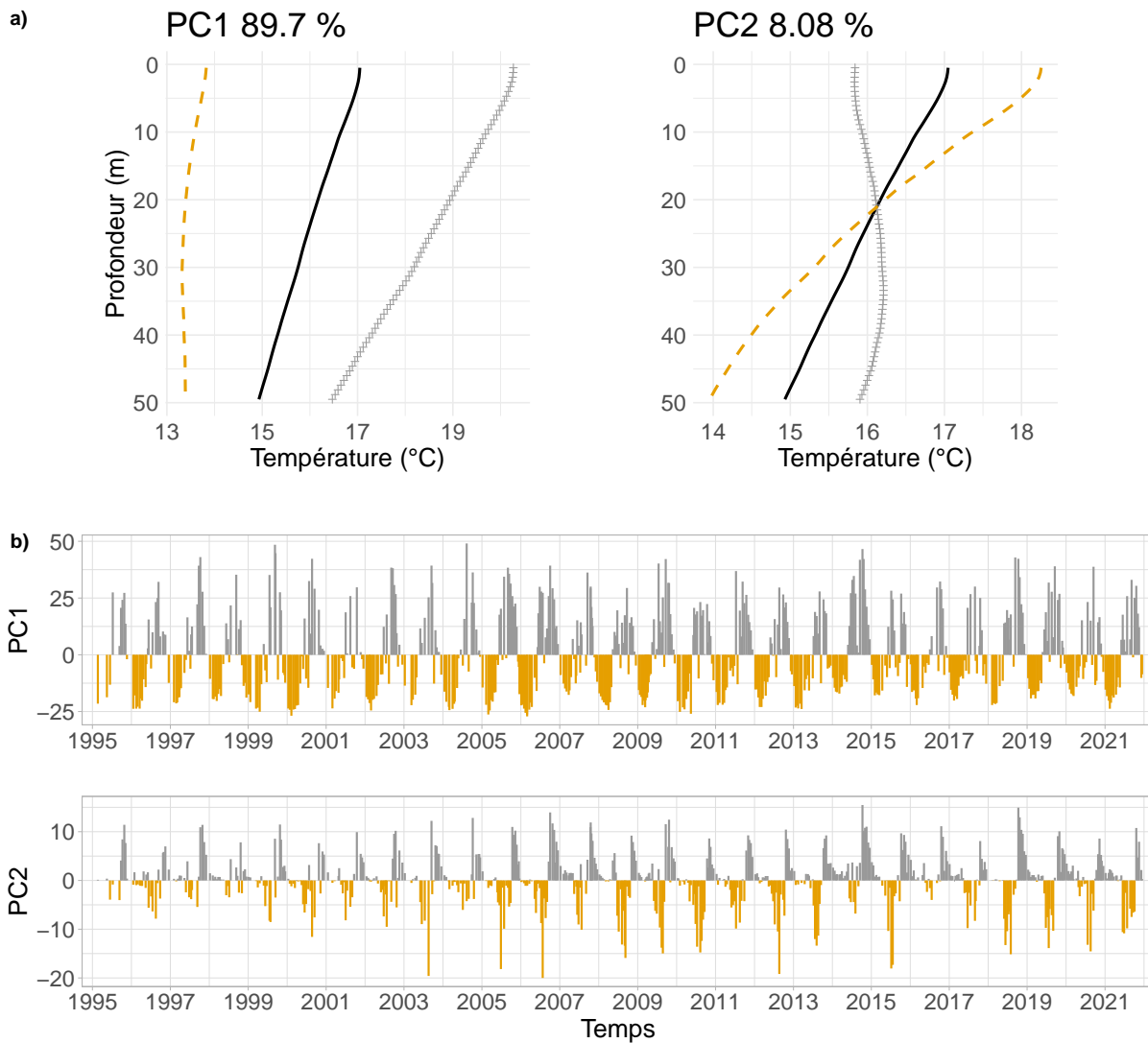


Figure 9 : a) Deux premières composantes principales de l'ACP fonctionnelle des profils de température à Marseille, exprimées comme perturbation de la moyenne. La courbe noire correspond au profil moyen. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative. b) Séries temporelles des scores des deux premières composantes principales.

Une manière de comparer les années entre-elles est de réaliser une ACP fonctionnelle sur les unités annuelles de chaque composante principale. Les séries temporelles des scores des PC1 et PC2 de l'ACP fonctionnelle précédente sont utilisées. Les ACP fonctionnelles sont réalisées séparément pour les deux

composantes. Pour chaque ACP, une fonction correspond à l'évolution du score de la composante sur une année. Les dates d'observations correspondent aux mêmes dates d'échantillonnage des profils CTD. Pour avoir des observations régulières, une moyenne mensuelle des scores pour chaque année est réalisée, ce qui représente donc 12 observations par an. Pour plusieurs années, certains mois ne comportent aucune observation, nous utilisons alors le paquet *fdapace* qui permet de combler ces lacunes pour conserver le plus grand nombre d'années dans la comparaison

La Figure 10 présente les résultats de l'ACP fonctionnelle sur la deuxième composante principale des profils verticaux. Rappelons qu'elle représente l'évolution de la stratification de la colonne d'eau au cours d'une année. La courbe moyenne en noire indique qu'on observe une forte stratification de la colonne d'eau en été, exprimée par des valeurs négatives du PC2. Au contraire, à la fin de l'année, la température de la colonne d'eau est plutôt homogène puisque les valeurs du PC2 sont positives. Les deux premières composantes résument 88,63% des variations observées entre chaque année, de 1994 à 2021. La première composante principale représente 73,71% de ces variations. Elle illustre une différence dans la magnitude, en même temps qu'un décalage dans le temps de la stratification. La deuxième composante principale, qui constitue 14,92% des variations, rend compte d'un effet moyen. La visualisation des scores de l'ACP sur deux dimensions montre que les premières années de la série sont plutôt regroupées dans le cadran inférieur gauche alors que les années les plus récentes se situent plutôt dans le cadran supérieur droit.

3.2 Variations dans l'espace

Nous cherchons maintenant à distinguer les stations en fonction de l'évolution de chaque variable au cours de l'année. Nous nous intéressons ici aussi, à des unités annuelles.

3.2.1 Abondance des groupes phytoplanctoniques

Pour savoir si les stations se distinguent en terme d'abondance de chaque groupe de phytoplancton au cours de l'année, nous nous intéressons tout d'abord aux données d'abondance.

Pour chaque groupe de phytoplancton toutes stations confondues, une ACP fonctionnelle est réalisée sur les données d'abondance. Il y a 5 groupes, nous allons donc effectuer 5 ACP différentes. Ici, une fonction correspond à l'évolution de l'abondance d'un groupe à une station, sur un an. Pour avoir des observations régulières, une moyenne mensuelle des données échantillonnées est réalisée pour chaque année. Il y a donc 12 points par an. Pour pouvoir conserver le plus grand nombre d'années dans la comparaison, le paquet *fdapace* est utilisé. Les deux premières composantes des ACP sont présentées en Annexe A

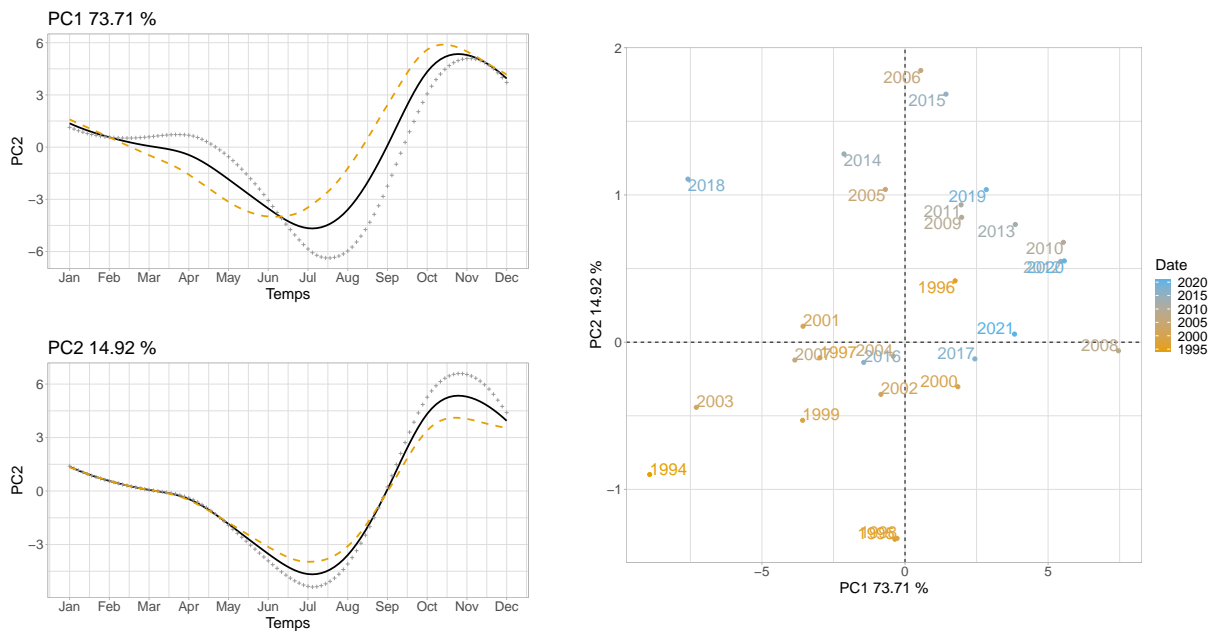


Figure 10 : (gauche) Deux premières composantes principales de l'ACP fonctionnelle de la PC2 des profils verticaux de température à Marseille, exprimées comme perturbation de la moyenne. La courbe noire correspond au profil moyen. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative. (droite) Score des deux premières composantes principales.

sous forme de perturbations de la moyenne. L'effet de toutes les premières composantes principales est un effet moyen où des valeurs positives impliquent une augmentation et des valeurs négatives une diminution. L'inertie des deuxièmes composantes principales des ACP sur les abondances varie entre 8,7% à 25,8%. Ces deuxièmes composantes principales sont conservées pour l'AFD.

Nous obtenons un nouveau tableau d'observations. Chaque observation correspond à une année, avec une variable qualitative qui correspond au site et 10 variables quantitatives qui correspondent aux scores des deux premières composantes principales des ACP pour chaque groupe de phytoplancton.

Les résultats de l'AFD sur ce tableau est présenté en Figure 11. Les deux axes résument bien 100% des informations. Le premier est celui qui contribue le plus à la discrimination, à hauteur de 94,94%. Le deuxième axe a donc une contribution de 5,06%. Le premier axe permet une bonne discrimination des sites de Banyuls et Marseille. Le deuxième axe quant à lui, discrimine très bien le site de Villefranche des deux autres.

La Figure 12 correspond au graphique des corrélations des variables associées à l'AFD. Les PC1 correspondent à l'abondance moyenne de chaque groupe. Les PC2 sont illustrées sur les bords de la Figure 12. La PC1 des *Prochlorococcus* est celle qui contribue le plus à l'axe 1. Cela signifie que l'abondance des *Prochlorococcus* est supérieure à Marseille de janvier à septembre (Annexe A). Les variables qui con-

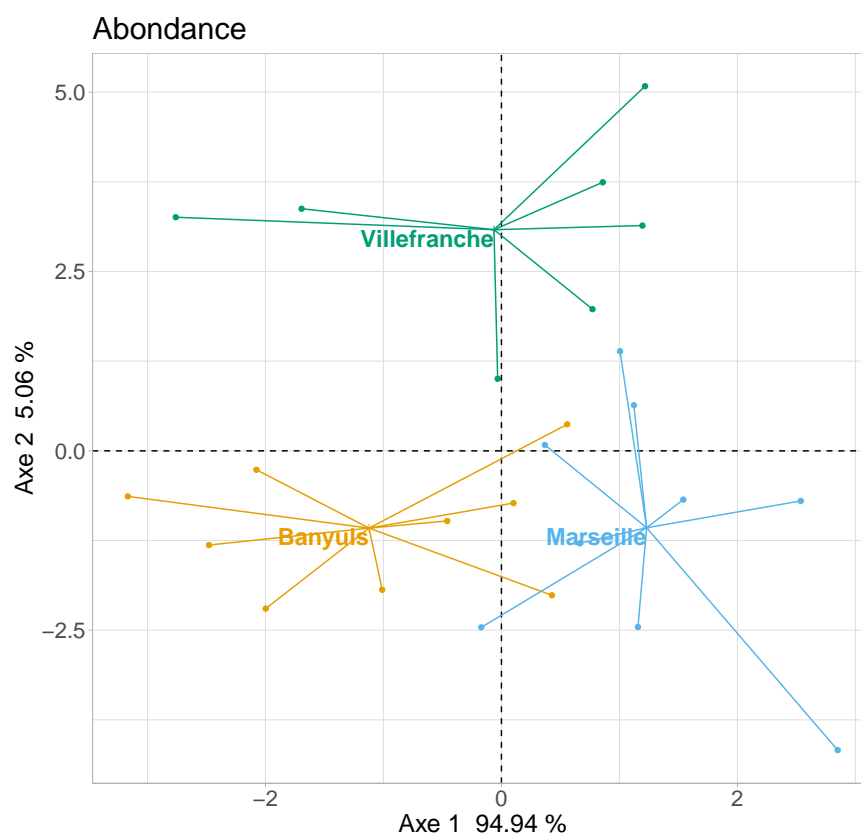


Figure 11 : Graphique des individus de l'AFD des unités annuelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.

tribuent le plus à l'axe 2 sont les PC2 des nano-eucaryotes et des *Prochlorococcus*. Cela indique d'une part, qu'à Villefranche l'abondance des nano-eucaryotes atteint un maximum plus importante de janvier à mai et diminue fortement le reste de l'année. À Banyuls et à Marseille, ce pic est moins important et l'abondance diminue moins le reste de l'année. D'autre part, à Villefranche l'abondance des *Prochlorococcus* est moins importante en début d'année, mais diminue moins en été qu'aux deux autres stations.

3.2.2 Diffusion lumineuse des groupes phytoplanctoniques

Nous voulons voir si les sites se distinguent par des populations présentant des organismes de taille différente.

Comme pour l'abondance, une ACP est réalisée pour chaque groupe de phytoplancton toutes stations confondues, avec les données de diffusion lumineuse. Les deux premières composantes des ACP sont présentées en Annexe C sous forme de perturbations de la moyenne. L'effet de toutes les premières composantes principales est un effet moyen où des valeurs positives impliquent une augmentation et des valeurs négatives impliquent une diminution. L'inertie des deuxièmes composantes principales des ACP

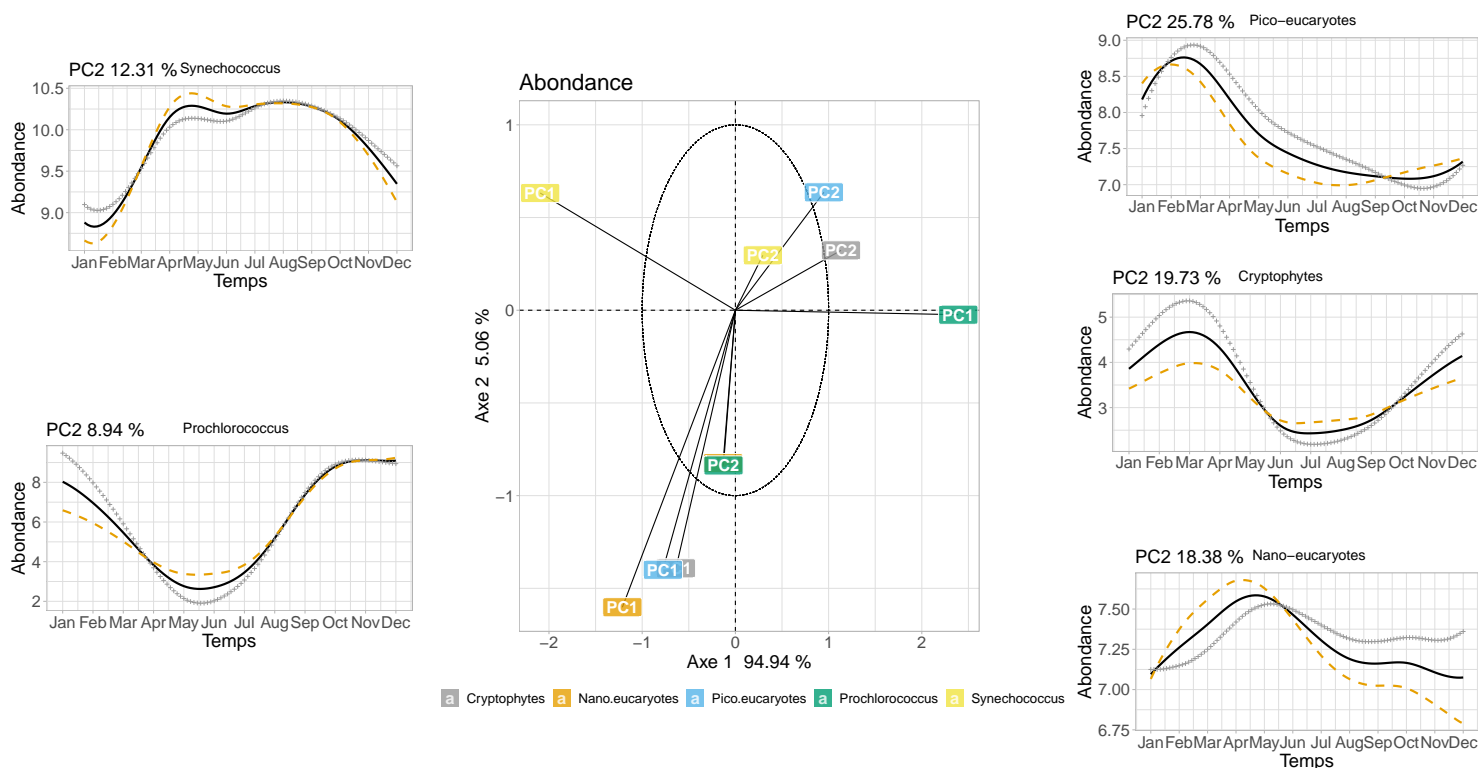


Figure 12 : (centre) Graphique des corrélations des variables de l'AFD des unités annuelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes. (côtés) Deuxièmes composantes principales des ACP fonctionnelles d'abondance du phytoplancton exprimées comme perturbation de la moyenne. La courbe noire correspond au profil moyen. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative.

est majoritairement inférieure à 10%. Nous choisissons de ne pas conserver ces deuxièmes composantes principales pour l'AFD.

On obtient un nouveau tableau d'observations. Chaque observation correspond à une année, avec une variable qualitative qui correspond au site et 5 variables quantitatives qui correspondent aux scores de la première composante principale des ACP pour chaque groupe du phytoplancton.

Les résultats de l'AFD sur ce tableau est présenté en Figure 11. Les deux axes résument bien 100% des informations. Le premier est celui qui contribue le plus à la discrimination, à hauteur de 90,18%. Le deuxième axe a donc une contribution de 9,82%. L'association des deux axes permet une bonne discrimination entre Banyuls et les deux autres sites. Cependant les sites de Marseille et Villefranche ne se distinguent pas très bien. En ce qui concerne la corrélation des variables, une forte SSC des nano-eucaryotes s'oppose à une forte SSC des *Synechococcus* et des pico-eucaryotes. La PC1 des *Prochlorococcus* est celle qui semble contribuer le plus à l'axe 2. À Banyuls les nano-eucaryotes ont des valeurs de SSC plus importantes toute l'année, mais des *Prochlorococcus*, des *Synechococcus* et des pico-eucaryotes avec des valeurs de SSC plus faibles toute l'année (Annexe C).

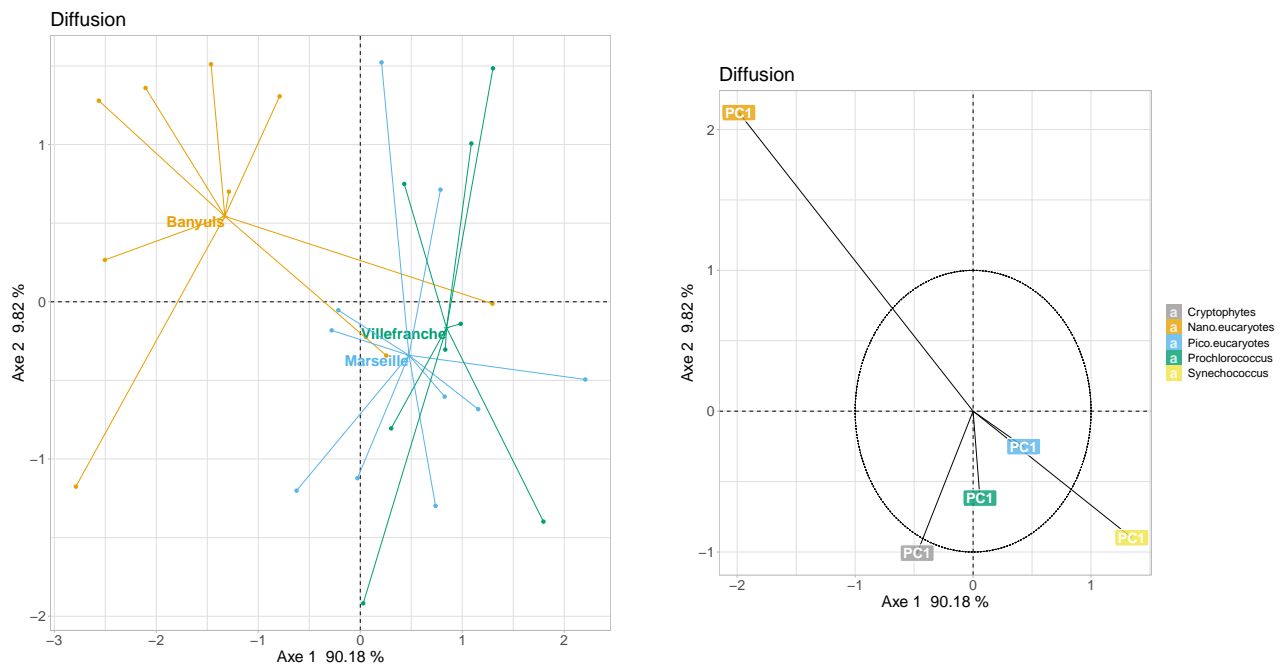


Figure 13 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'AFD des unités annuelles de diffusion lumineuse des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.

Nous remarquons également que la variation à l'intérieur des groupes est très élevée. Il semble donc pertinent de regarder l'évolution des PC1 pour chaque station. La Figure 14 présente l'évolution des PC1 au cours du temps pour les groupes de phytoplancton à chaque station. Les composantes principales ont été normalisées pour pouvoir être comparées entre-elles. À l'exception des Cryptophytes qui ont un comportement particulier, pour tous les autres groupes de phytoplancton, à chaque station nous observons que les valeurs de PC1 ont tendance à diminuer.

3.2.3 Nutriments

Enfin, nous cherchons à distinguer les sites par rapport à leur richesse en différents nutriments.

Cette fois, au lieu d'avoir 5 groupes de phytoplancton, nous nous intéressons à 5 nutriments NH_4 , NO_3 , NO_2 , PO_4 , $\text{Si}(\text{OH})_4$. Une ACP est réalisée pour chaque nutriment toutes stations confondues, avec leurs données de concentration. Les deux premières composantes des ACP sont présentées en Annexe B sous forme de perturbations de la moyenne. L'effet de toutes les premières composantes principales est un effet moyen où des valeurs positives impliquent une augmentation et des valeurs négatives impliquent une diminution. L'inertie des deuxièmes composantes principales des ACP est majoritairement inférieure à 15%. Nous choisissons donc de ne pas conserver ces deuxièmes composantes principales pour l'AFD.

Nous obtenons alors un nouveau tableau d'observations. Chaque observation correspond à une année,

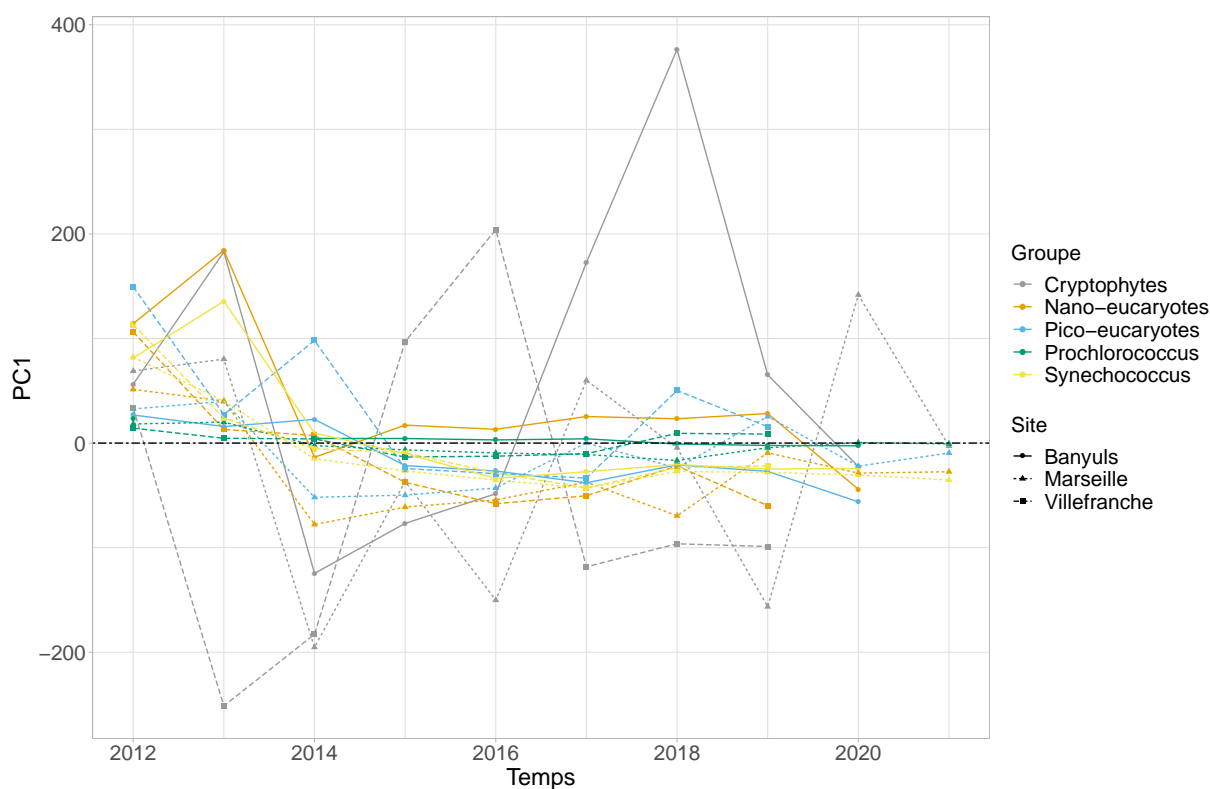


Figure 14 : Séries temporelles des premières composantes principales issues des ACP sur les données de diffusion lumineuse aux trois stations SOMLIT méditerranéennes.

avec une variable qualitative qui correspond au site et 5 variables quantitatives qui correspondent aux scores de la première composante principale des ACP pour chaque nutriment.

Les résultats de l'AFD sont présentés en Figure 15. Les deux axes résument bien 100% des informations. Le premier est celui qui contribue le plus à la discrimination, à hauteur de 76,18%. Le deuxième axe a donc une contribution de 23,82%. Le premier axe permet une bonne discrimination du site de Villefranche des deux autres. Le deuxième axe quant à lui, discrimine bien le site de Banyuls des deux autres. La corrélation des variables montre que de fortes concentrations en nitrate et en ammonium s'opposent à de fortes concentrations en phosphate et en nitrite. Le site de Banyuls est plutôt riche en nitrate et en ammonium et plus pauvre en phosphate et nitrite, à l'inverse de Villefranche. Le site de Marseille est particulièrement riche en silice, par rapport aux deux autres, mais avec des caractéristiques plus proches de Banyuls.

3.3 Co-variations

Nous voulons étudier comment la structuration des communautés est influencée par les conditions environnementales et notamment la richesse du milieu.

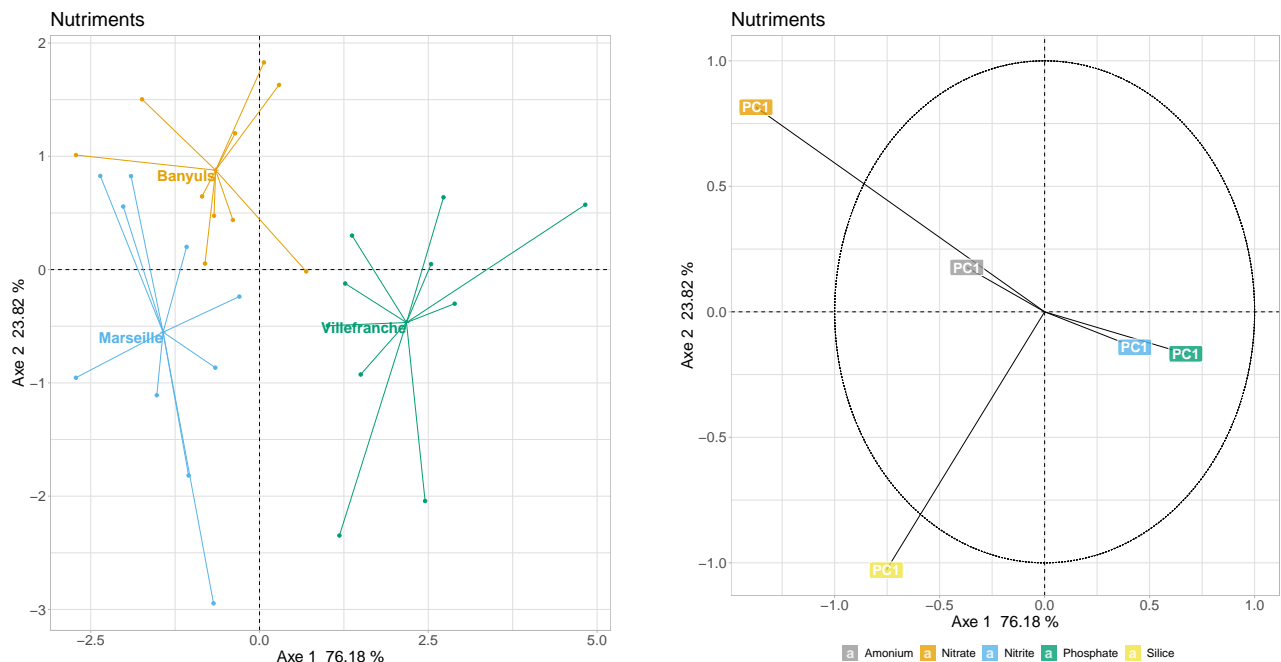


Figure 15 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'AFD des unités annuelles de concentration des nutriments pour les trois stations SOMLIT méditerranéennes.

3.3.1 Abondance et Nutriments

Dans un premier temps, nous nous intéressons au lien entre l'abondance des différents groupes de phytoplancton et la concentration de chaque nutriment.

Nous utilisons les scores des premières composantes principales des ACP sur les données d'abondance du phytoplancton et de concentrations en nutriments. Elles rendent toutes compte d'un effet moyen (Annexes A et B). Nous avons donc deux jeux de données, d'une part les données d'abondance et d'autre part les données environnementales. Chacun des jeux de données comporte 5 variables pour les 5 ACP respectives. Les individus pour chacun des tableaux sont les mêmes. Un individu correspond à une année, à un site.

L'ACC de ces deux tableaux produit les résultats présentés en Figure 16. Les deux premières corrélations canoniques pour les dimensions 1 et 2 s'élèvent respectivement à 0.853 et 0.717. Les trois autres corrélations correspondant aux dimensions 3, 4 et 5 sont inférieures à 0.5. Sur la Figure 16 seules les deux premières dimensions sont représentées. Le graphique de droite permet de comprendre les relations entre les variables. Plus elles sont proches, plus elles sont corrélées. Ce graphique met en évidence une forte corrélation entre l'abondance des *Prochlorococcus* et la concentration en phosphate. L'abondance des pico-eucaryotes semblent être plutôt corrélée à la concentration en nitrate. Le graphique des individus sépare assez bien les trois sites. Une opposition entre les sites de Villefranche et Banyuls apparaît, avec d'un côté un site

avec une concentration plus importante en PO_4 et une abondance plus importante des *Prochlorococcus* et de l'autre, un site avec une concentration plus élevée en NO_3 et une abondance des pico-eucaryotes plus importante.

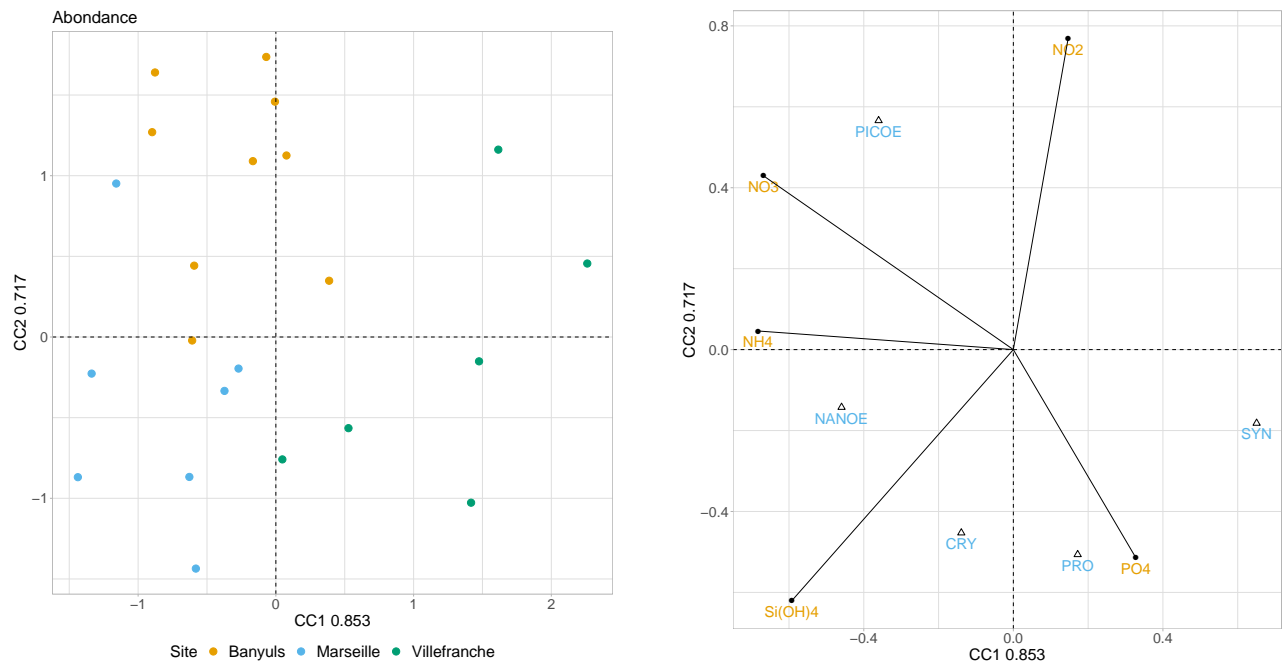


Figure 16 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'ACC des unités annuelles de concentration des nutriments et de l'abondance des groupes du phytoplancton pour les trois stations SOMLIT méditerranéennes. Le premier axe correspond à la dimension 1 avec une corrélation de 0.853. Le deuxième axe correspond à la dimension 2 avec une corrélation de 0.717. Sur le graphique de droite les nutriments sont représentés par des traits et les groupes de phytoplancton (*Prochlorococcus* (PRO), *Synechococcus* (SYN), pico-eucaryotes (PICOE), Cryptophytes (CRY) et nano-eucaryotes (NANOE)) par des triangles.

3.3.2 Diffusion lumineuse et Nutriments

Dans un second temps, nous nous intéressons au lien entre la taille des organismes des différents groupes de phytoplancton et la concentration de chaque nutriment.

Cette fois-ci, nous récupérons les scores des premières composantes principales des ACP sur les données de diffusion lumineuse du phytoplancton et de concentration des nutriments. Elles rendent toutes compte d'un effet moyen (Annexes B et C). Nous avons deux jeux de données, d'une part les données de diffusion SSC et d'autre part les données environnementales. Chacun des jeux de données comporte 5 variables pour les 5 ACP respectives. Les individus pour chacun des tableaux sont les mêmes. Un individu correspond à une année à un site.

L'ACC de ces deux tableaux produit les résultats présentés en Figure 17. Les deux premières corrélations

canoniques pour les dimensions 1 et 2 s'élèvent respectivement à 0.71 et 0.628. Les trois autres corrélations correspondant aux dimensions 3, 4 et 5 sont inférieures à 0.5. Sur la Figure 17 seules les deux premières dimensions sont représentées. Sur le graphique des corrélations, les segments correspondant aux variables environnementales sont pratiquement orthogonaux. Cela met en évidence des relations antagonistes des variables environnementales avec la diffusion lumineuse des groupes. Les nutriments riches en azote sont opposés au phosphate. Nous remarquons une forte corrélation entre la diffusion lumineuse des pico-eucaryotes et la concentration en phosphate. La diffusion lumineuse des Cryptophytes serait très liée à la concentration en NH_4 . Les *Synechococcus* sont très proches de l'origine. Cela signifie qu'elles ne sont pas bien représentées par ces deux dimensions. L'exploration des dimensions suivantes, permet de noter qu'elles ne sont pas mieux représentées. La SSC des *Synechococcus* ne semble donc corrélée à aucun des nutriments.

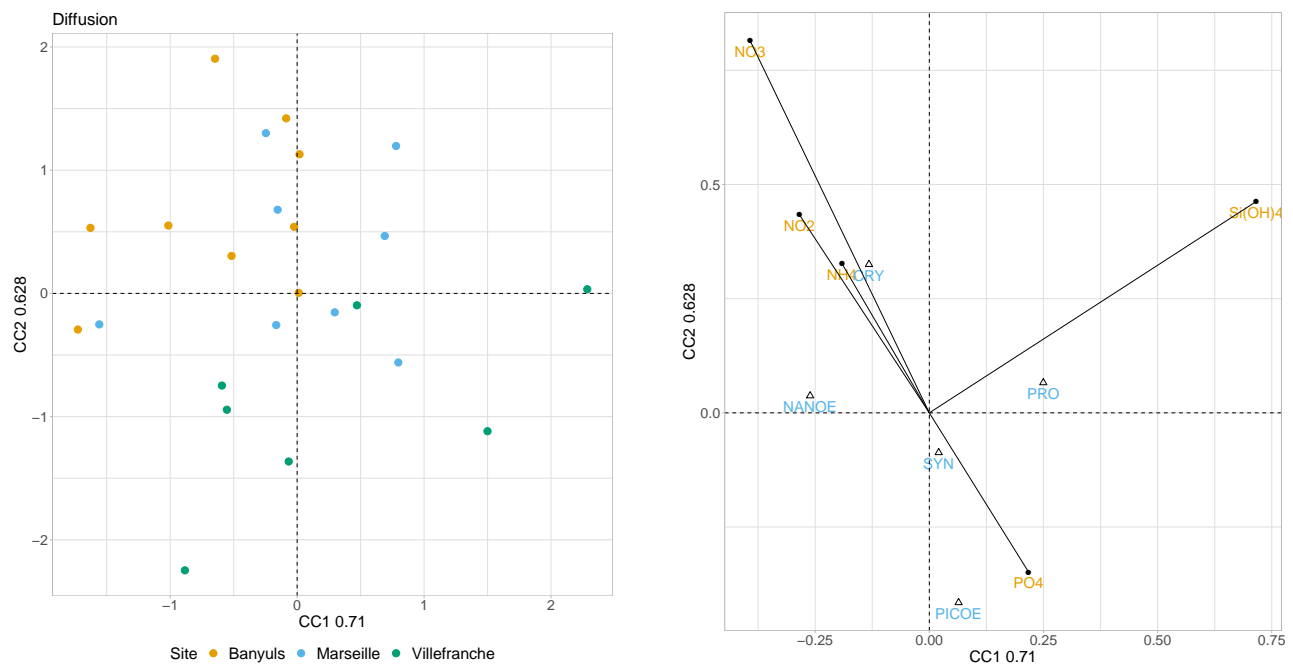


Figure 17 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'ACC des unités annuelles de concentration des nutriments et de la diffusion lumineuse des groupes du phytoplancton pour les trois stations SOMLIT méditerranéennes. Le premier axe correspond à la dimension 1 avec une corrélation de 0.71. Le deuxième axe correspond à la dimension 2 avec une corrélation de 0.628. Le deuxième axe correspond à la dimension 2 avec une corrélation de 0.717. Sur le graphique de droite les nutriments sont représentés par des traits et les groupes de phytoplancton (*Prochlorococcus* (PRO), *Synechococcus* (SYN), pico-eucaryotes (PICOE), Cryptophytes (CRY) et nano-eucaryotes (NANOE)) par des triangles.

4 DISCUSSION

Nous avons choisi de nous intéresser aux données collectées par le SNO SOMLIT pour les 3 stations méditerranéennes, Banyuls, Marseille et Villefranche. L'objectif de ce stage consiste à employer des méthodes d'analyses fonctionnelle et multivariée, pour (i) mettre en évidence les similitudes et dissimilitudes dans la structuration des communautés planctoniques, en tenant en compte des conditions environnementales de chaque station et (ii) caractériser la dynamique temporelle de ces communautés dans un contexte de changement global.

Dans un contexte de réchauffement climatique, la variable environnementale qui présente un intérêt évident est la température. Les profils verticaux de température pouvant être représentés par des courbes, l'ACP fonctionnelle est un très bon outils qui nous permet de comparer la forme de tous les profils échantillonnés à Marseille de 1994 à 2021. Cette méthode permet d'identifier les différences majeurs observées entre tous les profils. La première composante principale de l'ACP nous informe que presque 90% des différences sont dues à une variation de la température moyenne de la colonne d'eau. La deuxième composante principale nous informe que parmi les 10% restant, 8% des différences sont dues à la stratification de la colonne d'eau. Lorsque l'on visualise ces informations en fonction du temps, un motif saisonnier apparaît clairement pour les deux composantes principales. Les variations de la température moyenne de la colonne d'eau suivent une saisonnalité annuelle avec un refroidissement de la colonne d'eau en hiver et au printemps et un réchauffement de la colonne d'eau en été et en automne. Plus on avance dans le temps et plus la période où la température est supérieure à la moyenne dure longtemps, avec des températures de plus en plus chaudes. En hiver, l'eau a tendance à être de moins en moins froide. Nous observons donc bien une tendance qui va dans le sens du réchauffement climatique. La stratification de la colonne d'eau suit également une saisonnalité annuelle. Si la majeure partie de l'année la colonne d'eau est plutôt homogène, la stratification est plus importante pendant l'été. Nous mettons ainsi en évidence la formation de la thermocline saisonnière présente en Méditerranée (Millot, 1990). Cependant, la deuxième ACP réalisée sur les valeurs des deuxième composantes principales exprimées en fonction du temps nous permet de voir que les dernières années, la mise en place de la thermocline est retardée par rapport à la moyenne des années observées. Ce retard est associé à une stratification plus importante de la colonne d'eau. À l'inverse plus nous retournons dans le passé et plus nous observons la mise en place d'une thermocline moins marquée mais plus tôt dans l'année. De plus, en hiver la température en surface atteint des valeurs plus basses qu'entre 20 à 40m des profondeurs. Ces observations sont en accord avec les résultats obtenus par Rivetti *et al.* (2017) avec la modélisation de l'évolution de la thermocline de 1945 à 2011 à la station

DYFAMED en mer Ligure.

L'étude de la dynamique annuelle de concentration en nutriments à chaque station en utilisant les résultats des ACP couplés à une AFD permet une bonne discrimination des trois stations. Villefranche est un site particulièrement pauvre en nitrate et plus riche en phosphate en comparaison à Banyuls. Marseille semble présenter des caractéristiques plus proche de Banyuls, mais se distingue des deux sites par une concentration plus importante en silice. La mer Méditerranée est de nature oligotrophe. Cependant, à la différence de Villefranche, Marseille et Banyuls se situent dans des régions sous influence du Rhône et bénéficient donc des apports du fleuve. Le courant Nord associé aux vents, déplace les masses d'eau vers l'ouest (Pairaud *et al.*, 2011) et permet à Banyuls de bénéficier d'autant plus de ces apports.

La dynamique annuelle des groupes de phytoplancton permet également de bien discriminer les trois sites. Ici encore, Villefranche se distingue beaucoup des autres sites. D'une part, l'abondance des pico-eucaryotes, nano-eucaryotes et Cryptophytes y est moins importante qu'à Banyuls et Marseille. D'autre part, les variations intra-annuelles des *Prochlorococcus* et des nano-eucaryotes sont celles qui permettent de distinguer le mieux Villefranche des deux autres stations. La distinction entre Marseille et Banyuls se fait surtout par une abondance supérieure des *Prochlorococcus* à Marseille.

L'AFD réalisée avec les valeurs de diffusion lumineuse permet de séparer uniquement Banyuls des deux autres sites. À Banyuls, les nano-eucaryotes seraient plus grosses et les *Synechococcus* plus petites qu'aux deux autres sites. Néanmoins, l'inertie intra-classe est élevée. Cela signifie qu'il existe des variations non négligeables au sein même d'un site. En regardant l'évolution des composantes principales à chaque site, nous remarquons qu'à l'exception des Cryptophytes, elles ont tendance à diminuer. Nous pouvons en déduire qu'à tous les sites, hormis les Cryptophytes, les organismes phytoplanctoniques ont tendance à être de plus petite taille. De telles tendances ont déjà été mises en évidence chez le microplancton dans le Golfe de Naples (Ribera d'Alcalà *et al.*, 2004).

La mise en relation des variables biologiques et environnementales par l'ACC permet de comprendre en partie les résultats obtenus par les AFD. En effet, l'ACC met en évidence une corrélation positive entre l'abondance des pico-eucaryotes et la concentration en nitrate. Si l'abondance des pico-eucaryotes présente une corrélation négative avec les phosphates, les *Prochlorococcus* y sont positivement corrélés. L'abondance plus faible des pico-eucaryotes à Villefranche pourrait donc s'expliquer par la concentration plus faible en nitrate à cette station. Bien que les pico-eucaryotes soient des producteurs primaires avanta-gés dans un milieu mésotrophe, les *Prochlorococcus* sont largement dominant dans des conditions plus oligotrophes (Zubkov *et al.*, 2000; Pan, 2021). Nos résultats relatifs à l'abondance des pico-eucaryotes et des *Prochlorococcus* aux trois stations, Villefranche, Marseille et Banyuls, de la plus oligotrophe à la

moins oligotrophe, semblent donc pertinents.

En ce qui concerne les valeurs de diffusion lumineuse, considérées comme un indicatrices de la structure des cellules, l'ACC met en relief une forte opposition des effets des nutriments azotés et du phosphate. Il semble que les Cryptophytes présentent des cellules plus grosses lorsque le milieu est plus riche en ammonium ou qu'il est plus pauvre en phosphate. Et inversement pour les pico-eucaryotes. L'affinité particulière d'un groupe à l'azote ou au phosphate pourrait être expliquée par le fait qu'un nutriment soit plus limitant que l'autre. Étant donné qu'une petite taille de cellule confère un avantage lors de l'absorption des nutriments grâce à un plus grand rapport surface volume (Le Quéré *et al.*, 2005), il semble logique que la taille des cellules soit positivement corrélée à la concentration en nutriments. Cependant, si l'on associe les résultats de l'ACC et l'évolution des PC1 de diffusion lumineuse au cours du temps, il semble que cette relation ne soit pas celle qui prédomine pour les groupes autre que les Cryptophytes. Il faudrait explorer la corrélation de la diffusion lumineuse avec la température pour mettre en évidence si le réchauffement climatique pourrait être associé à de la diminution de la taille des cellules.

Pour confirmer ou explorer plus en détails les relations mises en évidence par l'ACC, il est envisageable de comparer deux courbes représentant des variables différentes à l'aide d'une ACC fonctionnelle (Ramsay & Silvermann, 2005).

Tous les groupes de phytoplancton n'étant pas bien représentés par l'ACC avec les variables relatives aux nutriments, une amélioration pourrait être faite en rajoutant d'autres variables environnementale avec lesquelles certains groupes présentent plus d'affinité, telles que la température et la salinité (Zubkov *et al.*, 2000). Cela pourrait permettre notamment d'expliquer pourquoi Banyuls se distingue autant des deux autres stations en terme de taille des nano-eucaryotes et des *Synechococcus*.

L'analyse des séries temporelles d'abondance met en évidence une saisonnalité marquée pour toutes les séries. Cette saisonnalité est différente d'un groupe à l'autre mais aussi d'une station à l'autre. Associées aux résultats des ACP sur les dynamiques annuelles de l'abondance de chaque groupe de phytoplancton, les séries temporelles mettent en évidence une variation intra-annuelle évidente au sein des groupes qui se répète globalement d'une année à l'autre. Ces variations sont cohérentes avec les précédentes études sur les variations saisonnières du phytoplancton en Méditerranée (Ribera d'Alcalà *et al.*, 2004; Bosc *et al.*, 2004). L'existence de variations intra-annuelles entre différents groupes de phytoplancton est bien référencée et peut être due aux affinités des groupes pour différentes conditions de température, d'éclairement, de stratification et de disponibilités en nutriments, qui suivent des cycles saisonniers (Levasseur *et al.*, 1984; Zubkov *et al.*, 2000; Romagnan *et al.*, 2015). De plus, il a déjà été mis en évidence que les variations intra-annuelles de la biomasse de phytoplancton sont dépendantes du système et notamment

de sa nature plus ou moins oligotrophe (Sommer *et al.*, 2012).

Il est possible d'extraire la saisonnalité d'une série temporelle avec des méthodes de décomposition saisonnalité-tendance (Cleveland *et al.*, 1990; Dokumentov & Hyndman, 2021). Ces méthodes ont été testées durant le stage, mais elles nécessitent d'être ajustées. Néanmoins, ce type de méthode permet de visualiser l'interaction d'un cycle saisonnier d'un groupe de phytoplancton avec un autre (Annexe D). Cela permettrait de comprendre comment les groupes se succèdent et interagissent entre eux ou avec une variable environnementale, comme par exemple la stratification de la colonne d'eau.

Les données SOMLIT, issues d'un échantillonnage bimensuel d'un grand nombre de variables, se prêtent bien aux méthodes d'analyses fonctionnelles et multivariées. Les observations sur 10 ans pour les données biologiques, et plus pour les données hydrologiques, permettent de constituer des séries d'observations suffisamment longues pour être exploitables. L'uniformisation des protocoles et de la mise en base des données pour toutes les stations SOMLIT facilite grandement les comparaisons inter-sites. Néanmoins, ce jeu de données est complexe. Il présente plusieurs dimensions : temporelle, spatiale (observations site-dépendantes) et verticale (profils CTD). La dimension temporelle peut être étudiée en dissociant deux échelles, intra et inter-annuelles, mais aussi en les emboîtant. Une difficulté supplémentaire s'ajoute avec la faible fréquence d'échantillonnage, qui est pourtant relativement importante en comparaison aux données habituellement disponibles pour les systèmes marins.

Si le caractère ataxonomique des données de cytométrie ne permet pas une étude de la composition spécifique, celles-ci permettent quand même une caractérisation des communautés basée sur des traits fonctionnels (Litchman & Klausmeier, 2008; Le Quéré *et al.*, 2005). Pour élargir le spectre des organismes étudiés, il peut être envisagé de compléter les données de pico et nano-plancton avec les données relatives au micro-planctons du réseau PHYTOBS.

Le fonctionnement des communautés planctoniques et l'impact que le changement global pourrait avoir sur elles, ont été plusieurs fois documentés comme étant site-dépendant (Cloern & Jassby, 2010; Sommer *et al.*, 2012; Domis *et al.*, 2013). Pour avoir une meilleure compréhension de la manière dont fonctionne les écosystèmes marins et des changements qui peuvent être attendus, il semble pertinent d'étudier le plus grand nombre de systèmes possibles. Un autre des avantages des données SOMLIT et que ces analyses peuvent être étendues au reste des stations métropolitaines.

5 CONCLUSION

Les analyses des données du SOMLIT conduites durant ce stage ont permis : i) de mettre en évidence la tendance de réchauffement de la colonne d'eau, ii) le décalage de la stratification estivale iii) de distinguer les trois stations méditerranéennes en fonction de leur nature trophique (plus ou moins oligotrophe), iv) d'illustrer les différentes dynamiques temporelles des groupes fonctionnels de phytoplancton à chaque station, v) d'identifier une affinité de certains groupes à la présence de certains nutriments, vi) d'observer une tendance à la diminution de la taille des cellules du phytoplancton.

Cette étude peut être complétée en poussant l'analyse des séries temporelles et de ses composantes tendance et saisonnalité. De même, il existe d'autres méthodes d'analyses fonctionnelles à exploiter pour explorer les corrélations entre variables (ACC fonctionnelle, STATIS...). D'autre part, il semble judicieux d'explorer les relations entre les variables biologiques et d'autres variables environnementales comme les précipitations ou l'influence des vents, et des courants. Il est également possible d'élargir le spectre d'organismes aux bactéries hétérotrophes et au micro-plancton qui ne sont pas analysables par cytométrie en flux comme le microphytoplancton (trop peu abondant) échantillonné par le réseau PHYTOBS. Une caractérisation plus fine de la biodiversité par des méthodes omics, comme par exemple le metabarcoding, apporterait une information complémentaire sur le suivi de la diversité. Enfin, ces méthodes pourraient être appliquées à toutes les stations du réseau SOMLIT.

L'analyse fonctionnelle et multivariée des données du SOMLIT apporte des pistes à explorer et offre des perspectives intéressantes en vue de l'étude spatio-temporelle du fonctionnement et de la structuration des communautés de phytoplancton, avec une approche basée sur les traits de ses groupes fonctionnels, dans le contexte du changement global.

BIBLIOGRAPHIE

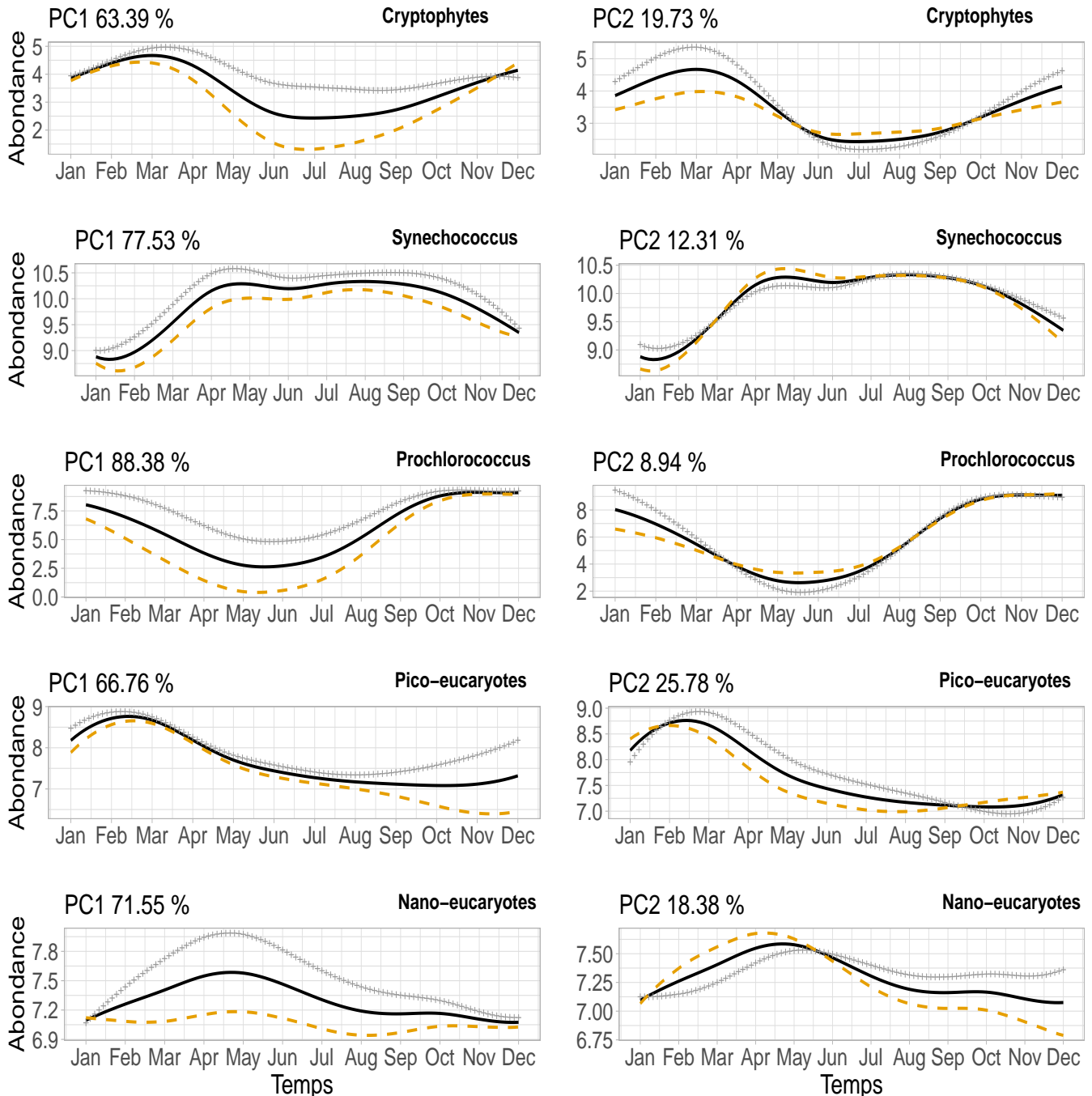
- Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* 10, 257–263.
- Bosc, E., Bricaud, A., Antoine, D. (2004). Seasonal and interannual variability in algal biomass and primary production in the mediterranean sea, as derived from 4 years of seawifs observations. *Global Biogeochemical Cycles* 18.
- Brunet, C., Casotti, R., Vantrepotte, V., Corato, F., Conversano, F. (2006). Picophytoplankton diversity and photoacclimation in the strait of sicily (mediterranean sea) in summer. i. mesoscale variations. *Aquatic Microbial Ecology* 44, 127–141.

- Cleveland, R. B., Cleveland, W. S., McRae, J. E. , Terpenning, I. (1990). Stl : a seasonal trend decomposition procedure based on loess. *Journal of Official Statistics* 6, 3–73.
- Cloern, J. E. , Jassby, A. D. (2010). Patterns and scales of phytoplankton variability in estuarine-coastal ecosystems. *Estuaries and Coasts* 33, 230–241.
- Dokumentov, A. , Hyndman, R. J. (2021). Str: Seasonal-trend decomposition using regression. *INFORMS Journal on Data Science* .
- Domis, L. N. D. S., Elser, J. J., Gsell, A. S., Huszar, V. L., Ibelings, B. W., Jeppesen, E., Kosten, S., Mooij, W. M., Roland, F., Sommer, U., Donk, E. V., Winder, M. , Lürling, M. (2013). Plankton dynamics under different climatic conditions in space and time. *Freshwater Biology* 58, 463–482.
- Dubelaar, G. B., Casotti, R., Tarran, G. A. , Biegala, I. C. (2007). Phytoplankton and their analysis by flow cytometry. *Flow cytometry with plant cells: Analysis of genes, chromosomes and genomes* pp. 287–322.
- Durrieu de Madron, X., Guieu, C., Sempéré, R., Conan, P., Cossa, D., D’Ortenzio, F. , *et al.* (2011). Marine ecosystems’ responses to climatic and anthropogenic forcings in the mediterranean. *Progress in Oceanography* 91, 97–166.
- Fan, J. , Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall.
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. , Falkowski, P. (1998). Primary production of the biosphere : Integrating terrestrial and oceanic components. *Science* 281, 237–240.
- González, I., Dejean, S., Martin, P. G., Baccini, A. , Martin, P. G. P. (2008). Cca: An r package to extend canonical correlation analysis. *Journal of Statistical Software* 23, 1–14.
- Josse, J. , Husson, F. (2012). Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistique* 153, 79–99.
- Le Quéré, C., Harrison, S. P., Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L. , *et al.* (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* 11.
- Levasseur, M., Therriault, J.-C. , Legendre, L. (1984). Hierarchical control of phytoplankton succession by physical factors. *Marine Ecology Progress Series* 19, 211–222.
- Litchman, E. , Klausmeier, C. A. (2008). Trait-based community ecology of phytoplankton. *Annual Review of Ecology, Evolution, and Systematics* 39, 615–639.
- Marie, D., Brussaard, C. P. D., Partensky, F. , Vaultot, D. (1999). Flow cytometric analysis of phytoplankton, bacteria and viruses. *Current protocols in cytometry* pp. 1–15.
- Marie, D., Rigaut-Jalabert, F. , Vaultot, D. (2014). An improved protocol for flow cytometry analysis of phytoplankton cultures and natural samples. *Cytometry Part A* 85, 962–968.
- McGill, B. J., Enquist, B. J., Weiher, E. , Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution* 21, 178–185.
- Millot, C. (1990). The gulf of lions’ hydrodynamics. *Continental Shelf Research* 10, 885–894.

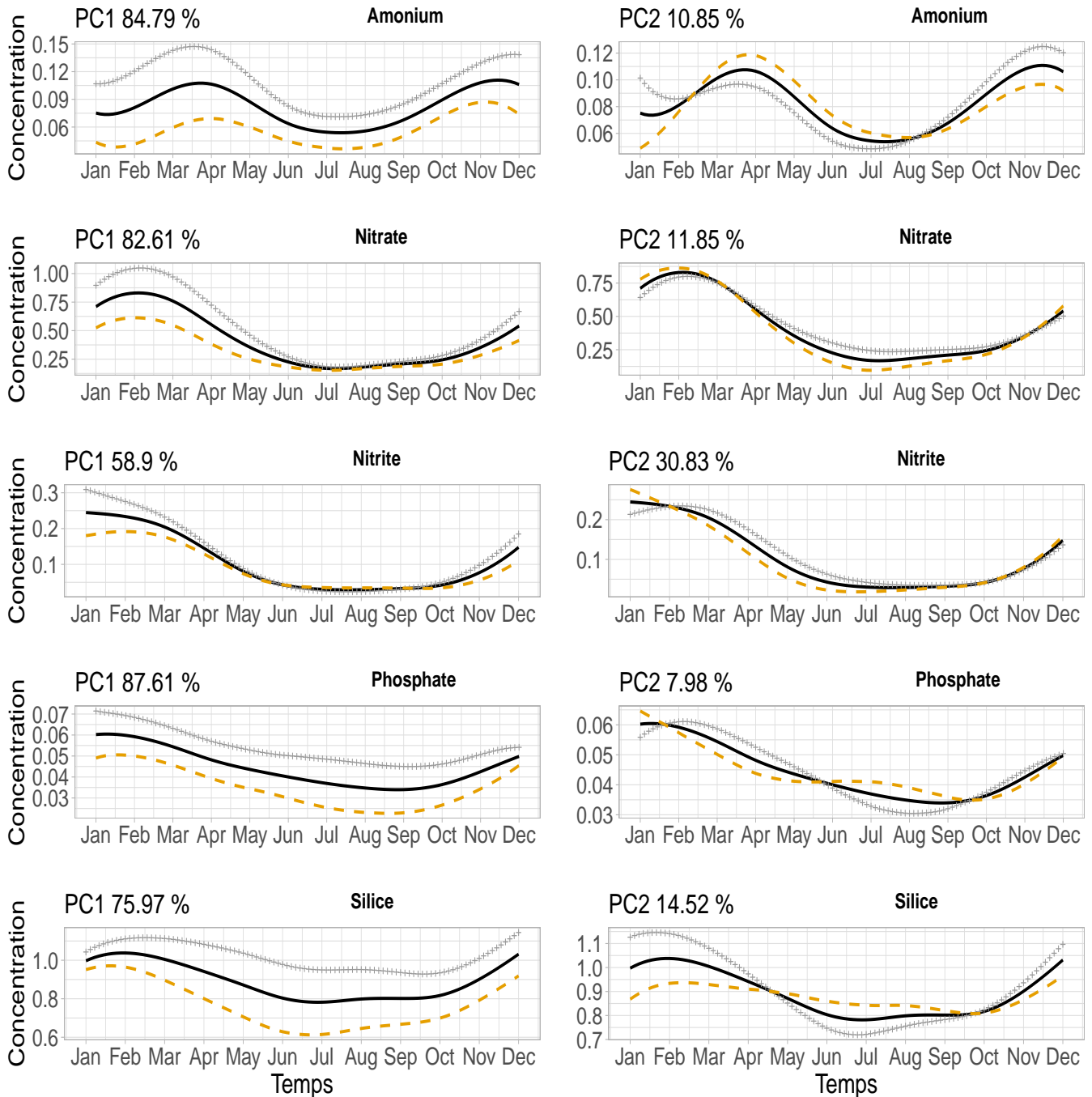
- Nerini, D., Manté, C. , Monestiez, P. (2021). Extending functional kriging when data are multivariate curves : some technical considerations and operational solutions. *Geostatistical functional data analysis* pp. 73–103.
- Pairaud, I. L., Gatti, J., Bensoussan, N., Verney, R. , Garreau, P. (2011). Hydrology and circulation in a coastal area off marseille: Validation of a nested 3d model with observations. *Journal of Marine Systems* 88, 20–33.
- Pan, J. (2021). Marine phytoplankton diversity and ecology. *Marine biology A functional approach to the oceans and their organisms* pp. 43–71.
- Pomeroy, L. R. (1974). The ocean’s food web, a changing paradigm. *BioScience* 24, 499–504.
- Ramsay, J. , Silvermann, B. (2005). *Functional data analysis*. Springer Series in Statistics.
- Ribera d’Alcalà, M., Conversano, F., Corato, F., Licandro, P., Mangoni, O., Marino, D., Mazzocchi, M. G., Modigh, M., Montresor, M., Nardella, M., Saggiomo, V., Sarno, D. , Zingone, A. (2004). Seasonal patterns in plankton communities in pluriannual time series at a coastal mediterranean site (gulf of naples): An attempt to discern recurrences and trends. *Scientia Marina* 68, 65–83.
- Rivetti, I., Boero, F., Frascchetti, S., Zambianchi, E. , Lionello, P. (2017). Anomalies of the upper water column in the mediterranean sea. *Global and Planetary Change* 151, 68–79.
- Romagnan, J. B., Legendre, L., Guidi, L., Jamet, J. L., Jamet, D., Mousseau, L., Pedrotti, M. L., Picheral, M., Gorsky, G., Sardet, C. , Stemann, L. (2015). Comprehensive model of annual plankton succession based on the whole-plankton time series approach. *PLoS ONE* 10.
- Shumway, R. H. , Stoffer, D. S. (2016). *Time series analysis and its applications*. quatrième edn. Springer Texts in Statistics.
- Simonoff, J. S. (1996). *Smoothing methods in statistics*. Springer series in Statistics.
- Sommer, U., Adrian, R., Domis, L. D. S., Elser, J. J., Gaedke, U., Ibelings, B., Jeppesen, E., Lürling, M., Molinero, J. C., Mooij, W. M., Donk, E. V. , Winder, M. (2012). Beyond the plankton ecology group (peg) model: Mechanisms driving plankton succession. *Annual Review of Ecology, Evolution, and Systematics* 43, 429–448.
- Trask, B. J., Engh, G. J. V. D. , Elgershuizen, J. H. B. W. (1982). Analysis of phytoplankton by flow cytometry. *Cytometry* 2.
- Våge, S. , Thingstad, T. F. (2015). Fractal hypothesis of the pelagic microbial ecosystem - can simple ecological principles lead to self-similar complexity in the pelagic microbial food web ?. *Frontiers in Microbiology* 6, 1–13.
- Zubkov, M. V., Sleigh, M. A., Burkill, P. H. , Leakey, R. J. G. (2000). Picoplankton community structure on the atlantic meridional transect: a comparison between seasons. *Progress in Oceanography* 45.

ANNEXES

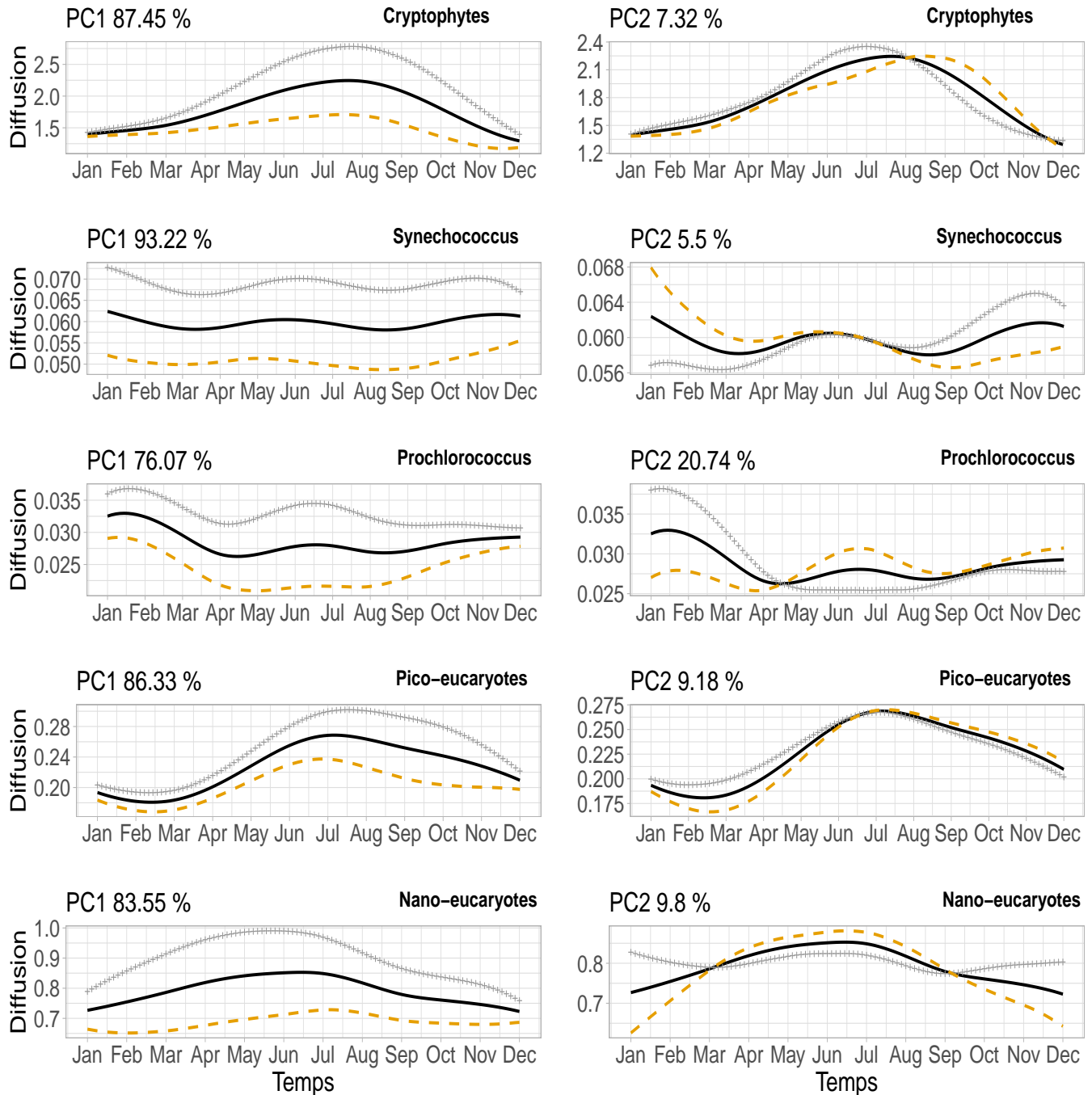
A. Deux premières composantes principales des ACP fonctionnelles des données d'abondance pour chaque groupe de phytoplancton des trois stations SOMLIT méditerranéennes, exprimées comme perturbation de la moyenne. Les courbes noires correspondent aux courbes moyennes. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative.



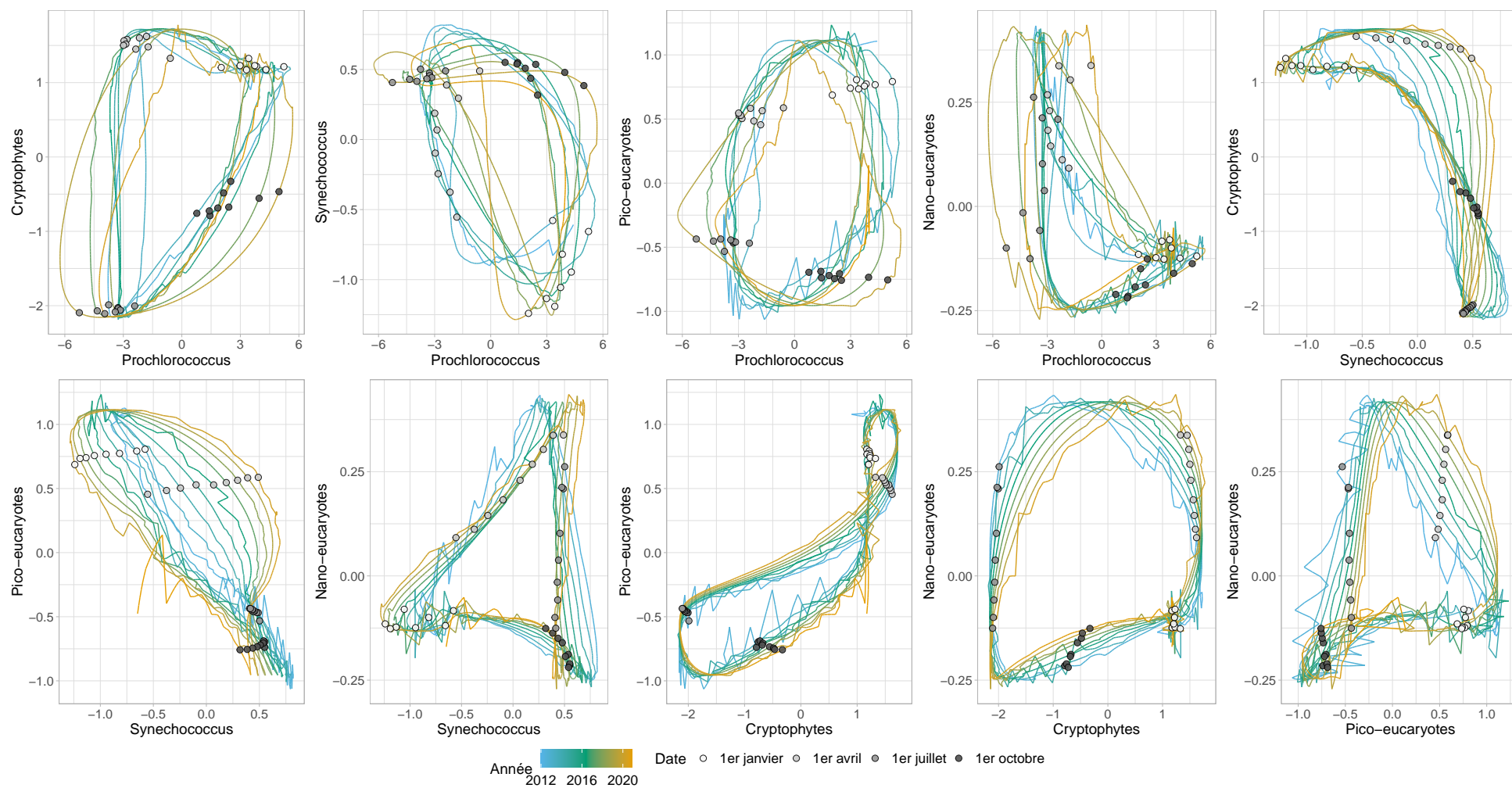
B. Deux premières composantes principales des ACP fonctionnelles des données de concentration pour chaque nutriment aux trois stations SOMLIT méditerranéennes, exprimées comme perturbation de la moyenne. Les courbes noires correspondent aux courbes moyennes. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative.



C. Deux premières composantes principales des ACP fonctionnelles des données de diffusion lumineuse pour chaque groupe de phytoplancton des trois stations SOMLIT méditerranéennes, exprimées comme perturbation de la moyenne. Les courbes noires correspondent aux courbes moyennes. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté l'effet d'une perturbation négative.



D. Diagrammes de phase des composantes saisonnières des séries temporelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.



RÉSUMÉ

L'accumulation des données grâce au développement d'une démarche systématique d'observation des milieux littoraux rend aujourd'hui possible l'étude de séries à long terme. Durant ce stage, nous nous intéressons aux assemblages phytoplanctoniques caractérisés par cytométrie en flux depuis 2012 aux trois stations SOMLIT méditerranéennes. La particularité de ce stage réside dans l'utilisation de méthodes statistiques multivariées et fonctionnelles. Cette approche présente l'avantage de pouvoir traiter les données temporelles et spatiales apparaissant sous forme de courbes échantillonnées. Nous avons ici recours à des méthodes de lissages non-paramétriques et d'analyse de séries temporelles pour mettre en évidence les tendances, dans le contexte du changement global. Des méthodes de classification sont utilisées pour discerner les stations en fonction de la structure de leur communauté et en fonction de leur richesse en nutriments. À l'aide de méthodes d'analyses multivariées, nous mettons en évidence les relations et entre les variables biologiques et les variables environnementales.

ABSTRACT

Thanks to consistent surveying of coastal ecosystems, we are now able to analyse pluri-annual time series of the data gathered over the years. During this study, we focus on the composition of phytoplankton communities characterised by flow cytometry since 2012 at the three Mediterranean SOMLIT sites. The specificity of this study lies in the use of multivariate analysis and functional data analysis. These methods enable us to process sampled data that can be viewed as curves expressed over time and over space. We use non-parametric smoothing methods and time series analysis to detect trends in the light of global warming. Classification methods are employed to discriminate sites based on the structure of their community and on nutrient availability. Multivariate statistics allows us to uncover correlations between biological variables and environmental variables.