

AIX-MARSEILLE UNIVERSITÉ

Institut OSU-Pythéas

Institut Méditerranéen d'Océanologie - MIO

MASTER SCIENCES DE LA MER

Parcours Océanographie Biologique et Écologie Marine

Titre du stage

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

SOMMAIRE

1	Introduction	1
2	Matériel & Méthodes	3
2.1	SOMLIT (Service d'Observation en Milieu LITToral)	3
2.1.1	Les stations	3
2.1.2	Prélèvements et protocoles	4
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	13
2.3.1	Lissage paramétrique	13
2.3.2	ACP fonctionnelle	14
2.3.3	Perturbation de la moyenne	14
2.4	Analyse Factorielle Discriminante	14
2.5	Analyse Canonique des Corrélations	15
3	Résultats	16
3.1	Variations dans le temps	16
3.1.1	Abondance	16
3.1.2	Température	18
3.2	Variations dans l'espace	21

3.2.1	Abondance	21
3.2.2	Diffusion lumineuse	23
3.2.3	Nutriments	25
3.3	Co-variations	26
3.3.1	Abondance et Nutriments	26
3.3.2	Diffusion lumineuse et Nutriments	27
4	Discussion	28
5	Conclusion	28
Bibliographie		28
Liste des Figures		29
Liste des Tableaux		32
Annexes		33

1 INTRODUCTION

Les assemblages microbiens planctoniques jouent un rôle fondamental pour le fonctionnement de l’Océan (Pomeroy, 1974, Azam, 1983). L’étude de leur composition, de leur distribution et de leur dynamique est une étape clé dans la compréhension du fonctionnement et de l’évolution de l’écosystème marin dans le contexte du changement global et des forçages anthropiques. Parmi ces assemblages microbiens, le phytoplancton joue un rôle de première importance. Bien que sa biomasse ne représente qu’environ 2 pourcents de la biomasse globale, il est responsable de près de la moitié de la production primaire de la planète (Field, 1998). Il représente également le premier maillon du réseau trophique.

Parmi les méthodes d’investigation du phytoplancton, la cytométrie en flux est une technologie qui s’impose de plus en plus car elle permet l’analyse à haute vitesse des cellules (jusqu’à plusieurs milliers par seconde). Les cellules sont entraînées dans l’instrument par un liquide vecteur qui les sépare, les aligne et les conduit vers une source lumineuse. Elles sont interceptées une par une par cette source lumineuse (en général un ou plusieurs faisceaux laser) et les propriétés optiques de diffusion de la lumière et d’émission de fluorescence de ces cellules sont enregistrées. Ces variables permettent de discriminer les cellules présentes en différents groupes, et d’en déterminer leurs abondances. L’analyse rapide permet d’envisager l’étude de la distribution des assemblages microbiens à une fréquence impossible par des méthodes plus conventionnelles (type microscopie) (Gasol and Moran, 2015).

L’observation à moyen et long terme de l’évolution de l’environnement est clairement reconnue comme une nécessité impérative si l’on veut comprendre, via l’étude des ”séries à long terme”, comment les écosystèmes marins réagissent à la fois aux contraintes naturelles de l’environnement et aux effets anthropiques. C’est le cas des milieux littoraux (lagunes, baies, estuaires et l’ensemble des côtes) particulièrement vulnérables en raison de la pression croissante des activités humaines et objets de nombreux conflits d’usage. Le développement d’une démarche systématique d’observation des milieux littoraux est issu d’initiatives locales, fondées, soit sur des opportunités, soit sur une tradition de culture scientifique propre à chaque Station/Laboratoire Marin (RESOMAR). La fédération de ces efforts a conduit à la mise en place du Service d’Observation en Milieu Littoral (SOMLIT ; <http://somlit.epoc.u-bordeaux1.fr/fr/>) en 1995, labellisé par l’INSU depuis 1996, et qui regroupe actuellement 12 Stations Marines.

Depuis plus de 10 ans, la cytométrie en flux est utilisée par le SNO SOMLIT pour suivre la dynamique temporelle du pico- (taille $< 2 \text{ }\mu\text{m}$) et du nanophytoplancton (taille comprise entre 2 et 20 μm) à ses différentes stations. Ce suivi est complémentaire du suivi taxonomique du phytoplancton de plus grande taille du réseau PHYTOBS. Si la cytométrie en flux est ataxonomique, elle 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 tailles des cellules : pico-, nano et microplancton (Siebirth et al, 1978).

Les objectifs de ce stage sont (i) d'étudier la dynamique temporelle des assemblages phytoplanctoniques caractérisés par cytométrie en flux depuis 2012, (ii) de mettre en lien cette dynamique avec les variables hydrologiques et physico-chimiques échantillonnées simultanément (température, salinité, concentrations en nutriments, pH, etc.), (iii) d'évaluer l'importance de ces variables sur la structure dynamique planctonique. L'approche statistique proposée repose sur l'utilisation de méthodes statistiques qui permettent de traiter de données temporelles et spatiales apparaissant sous forme de courbes échantillonnées (profils verticaux des CTD ou les séries temporelles des mesures effectuées aux stations des réseaux SOMLIT et PHYTOBS). Les analyses peuvent être réalisées en intégrant les différentes échelles emboitées (intra et inter annuelles) (Nerini et. al., 2021).

Un focus particulier sera apporté aux 3 stations Méditerranéennes car elles correspondent à des conditions environnementales contrastées, dans une mer reconnue comme un hot spot de biodiversité très sensible au changement global (Durieu de Madron et al., 2011). 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 Sola et Frioul (?). 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 (?). 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 (?). 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 (?).

Ce stage cherchera à identifier les ressemblances et les différences entre les stations, aussi bien en termes de structure des communautés (groupes présents/absents, abondances) qu'en termes de dynamique temporelle. Le sujet combinera les données biologiques relatives au phytoplancton collectées dans les SNO SOMLIT (paramètre pico-nano) et du réseau d'observation du phytoplancton PHYTOBS.

L'originalité de ce sujet réside dans l'utilisation de méthodes statistiques multivariées, fonctionnelles, pour extraire au mieux l'information contenue dans ce suivi pluriannuel bimensuel, imputer des données manquantes en utilisant cet aspect fonctionnel et mettre en évidence de manière objective les tendances, dans le contexte du changement global. C'est à notre connaissance la première fois que le jeu de données biologiques de cytométrie SOMLIT sera étudié dans son ensemble, couplé aux données du réseau PHYTOBS, environné par les données hydrologiques (variables physico-chimiques) environnementales associées, en concertation avec les différents responsables SOMLIT et PHYTOBS des différentes stations. Ce sujet est donc par définition interdisciplinaire et trans-réseaux.

2 MATÉRIEL & MÉTHODES

2.1 SOMLIT (Service d'Observation en Milieu LITToral)

Le Service d'Observation en Milieu Littoral (SOMLIT) est un Service National d'Observation (SNO) qui existe depuis 1996. L'objectif du SOMLIT est de '*caractériser l'évolution pluri-décennale des écosystèmes côtiers et littoraux et d'en déterminer les forçages climatiques et anthropiques*'. 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).

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.

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

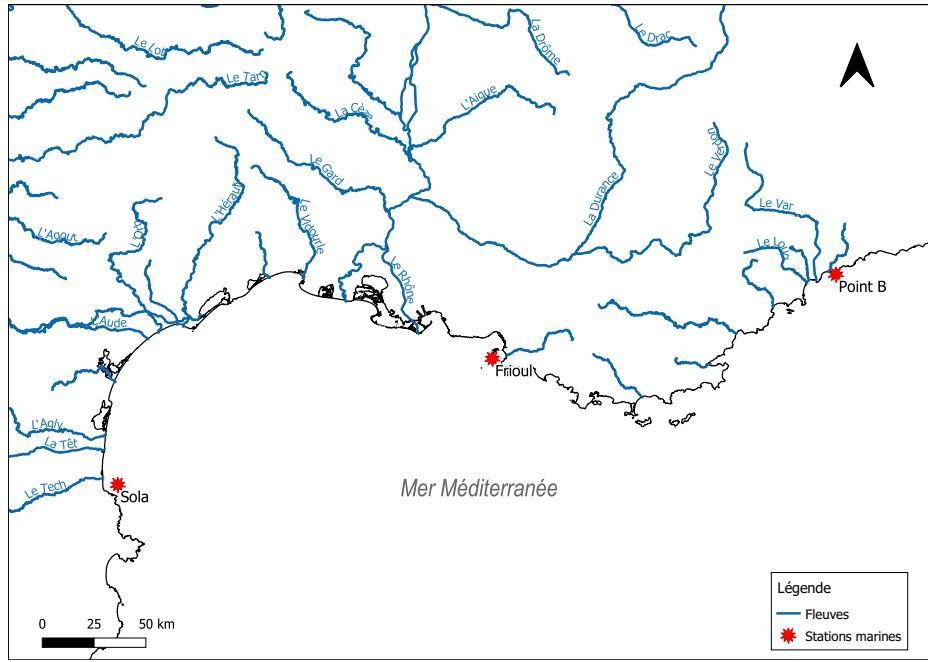


Figure 1 : Carte des stations SOMLIT méditerranéennes : Sola à Banyuls, Frioul à Marseille et Point B à Villefranche.

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. Ce 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.

2.1.2.1 Données hydrologiques

À chaque sortie en mer, l'eau est prélevée dans des bouteilles Niskin en surface (1m) 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'amonium (NH_4) par fluorimétrie. D'autre échantillonnages des bouteilles Niskin sont destinés à l'analyse des nutriments, nitrites (NO_3), nitrates (NO_2), phosphates (PO_4) et silicates (Si(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/>).

2.1.2.2 Profils verticaux

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.

2.1.2.3 Données biologiques

L'eau prélevée à la surface dans une bouteille Niskin est également échantillonnée pour les analyses du pico et nano-plancton (Marie et al., . 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 (?).

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 (?) et la diffusion de la lumière aux petits angles est en relation avec la taille de la particule (?).

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 on distingue les *Prochlorococcus*, les *Synechococcus*, les Cryptophytes, les pico-eucaryotes et les nano-eucaryotes (?).

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, les Dinoflagellés et le nano-plancton hétérotrophe (?). Les pico-eucaryotes autotrophes, d'une taille inférieur à $2\mu\text{m}$, présentent une forte affinité pour les nutriments et la lumière, due à un ratio surface/volume élevé (?). 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 , carbone (COP) et azote (NOP) organiques particulières, , matière en suspension (MES) et chlorophylle *a*. Le jeu de données biologiques (PICONANO) contient les données de cytométrie 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 de fluorescence verte moyenne. Les valeurs d'auto-fluorescence orange de *Synechococcus* et des Cryptophytes sont

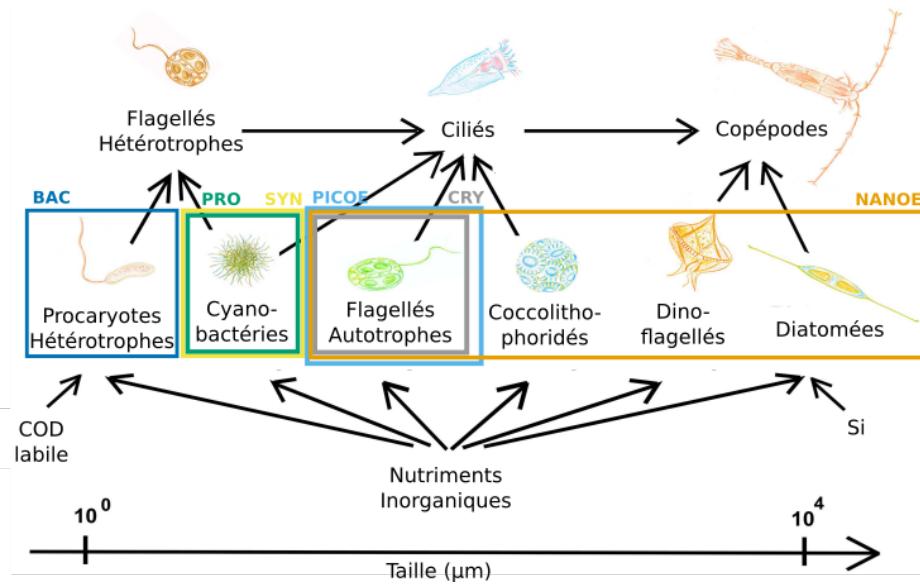


Figure 2 : Réseau trophique planctonique pélagique. Les groupes distingués par cytométrie en flux sont représentés par des cases qui englobent les organismes qu'ils peuvent contenir, en bleu foncé les bactéries hétérotrophes (BAC), en vert les *Prochlorococcus* (PRO), en jaune les *Synechococcus* (SYN)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

également renseignées. Le jeu de données CTD contient les profils verticaux de la surface au fond pour les variables température, salinité, chlorophylle *a* et éclairement.

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						

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.

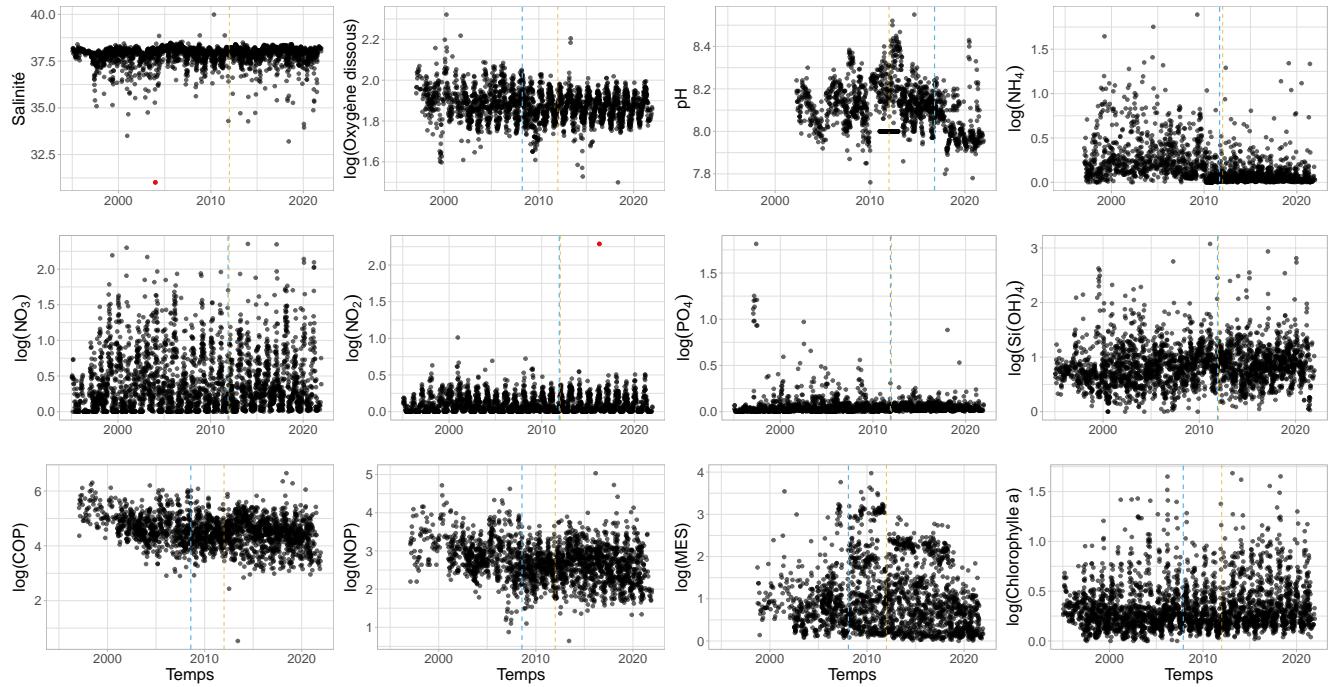


Figure 3 : Séries temporelles des données hydrologiques de surface pour les trois stations SOMLIT méditerranéennes confondues. La ligne pointillée bleue correspond à la mise en place du protocole national pour la variable présentée. La ligne pointillée orange correspond au début du jeu de données PICONANO.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

Dans le cadre de ce stage, nous nous sommes focalisés sur le phytoplancton étudié. 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-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

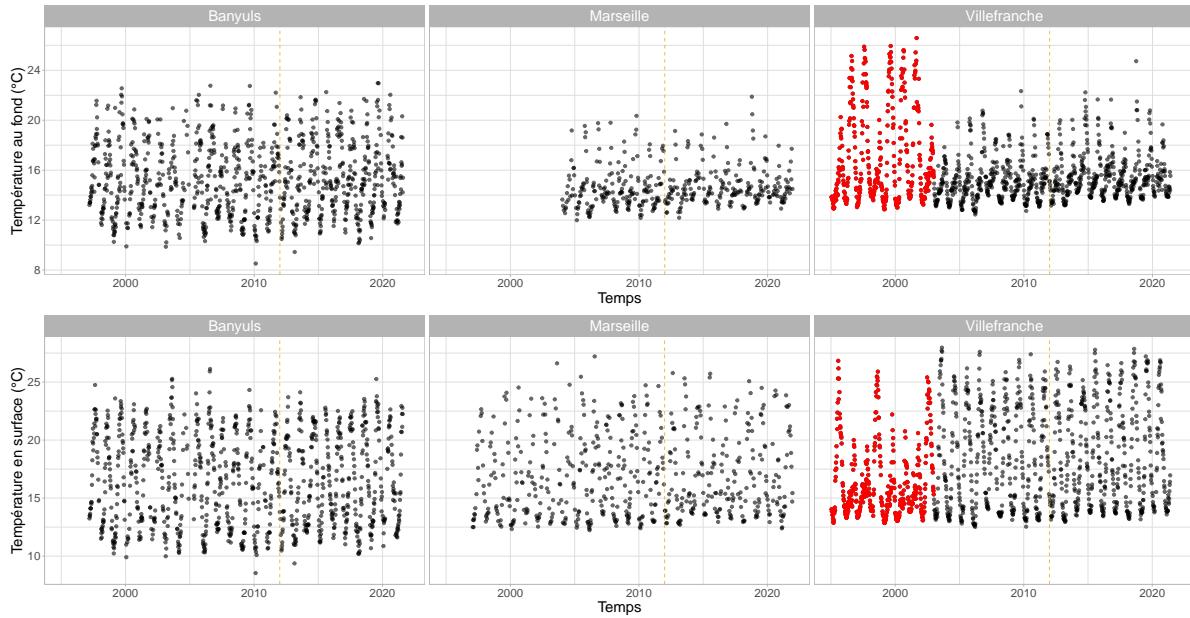


Figure 4 : Séries temporelles des données de température en surface et au fond pour les trois stations méditerranéennes. La ligne pointillée orange correspond au début du jeu de données PICONANO. Les éventuelles données aberrantes sont représentées en rouge.

toutes les autres, les données paraissent fiables puisque l'on identifie facilement les différents groupes.

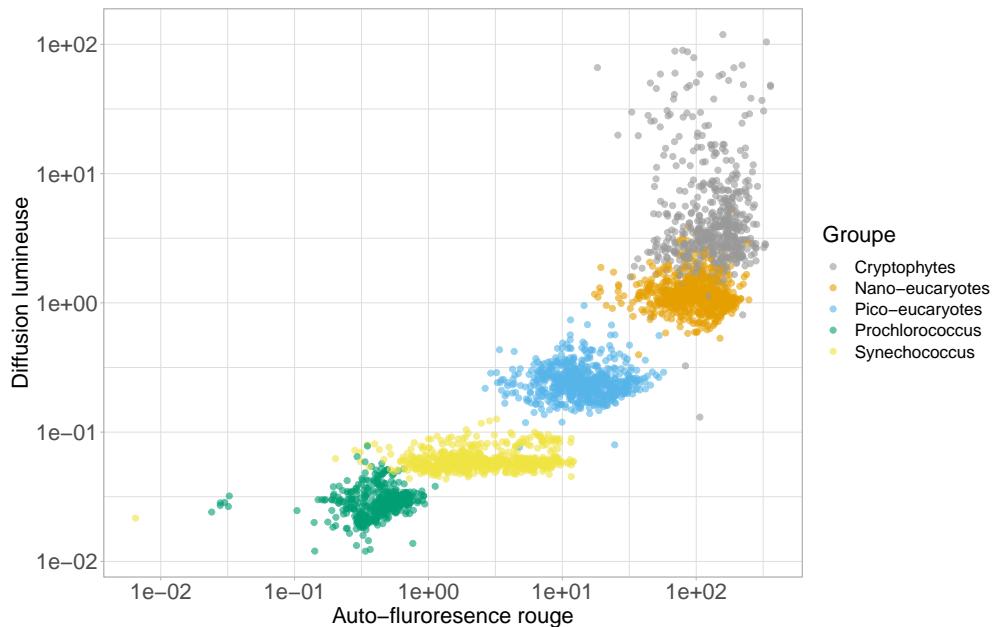


Figure 5 : Auto-fluorescence rouge moyenne *versus* diffusion lumineuse à 90° moyenne de chaque groupe de phytoplancton pour chaque prélèvement pour les trois stations SOMLIT méditerranéennes.

En ce qui concerne les données PICONANO, toutes les variables semblent exploitables et pourraient potentiellement apporter des informations pertinentes. Néanmoins, durant ce stage l'étude est focalisée

sur les variables d'abondance et de diffusion lumineuse (SSC), cette dernière étant indicatrice de la taille moyenne des organismes.

2.2 Analyse de séries temporelles

TEXTE

2.2.1 Lissage non-paramétrique

On souhaite explorer les séries temporelles pour identifier une structure et déceler une éventuelle tendance. Pour cela, dans un premier temps, on va réaliser un lissage sur chaque variable d'intérêt et pour chaque station. On choisit de réaliser un lissage non paramétrique puisque 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 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 (?).

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

Soit la matrice design X_x :

$$X_x = \begin{pmatrix} 1 & x - x_1 & \dots & (x - x_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x - x_n & \dots & (x - x_n)^p \end{pmatrix}$$

La matrice de poids W_x :

$$W_x = h^{-1} \text{diag} \left[K\left(\frac{x - x_1}{h}\right), \dots, K\left(\frac{x - x_n}{h}\right) \right]$$

On peut obtenir les coefficients $\hat{\beta}$:

$$\hat{\beta} = (X'_x W_x X_x)^{-1} X'_x W_x y$$

On fait le choix d'utiliser un noyau $K()$ gaussien, qui est non-borné et qui permet d'avoir une variance conditionnelle finie (?). Le choix de la fenêtre de lissage h est fait en testant plusieurs valeurs. On conserve 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 ?. 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 une éventuelle saisonnalité dans les séries temporelles, on s'intéresse 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.

Pour tout échantillon de séries temporelles x_1, \dots, x_n avec n impaire on peut écrire :

$$x_t = a_0 + \sum_{j=1}^{(n-1)/2} [a_j \cos(2\pi t j/n) + b_j \sin(2\pi t j/n)]$$

Le périodogramme échelonné peut être défini de la manière suivante :

$$P(j/n) = a_j^2 + b_j^2$$

Le périodogramme indique quelles composantes de fréquences ont une grande magnitude et lesquelles ont une faible magnitude. De grandes valeurs de $P(j/n)$ indiquent donc quelles fréquences $\omega_j = j/n$ sont prédominantes dans la série. Les petites valeurs de $P(j/n)$ 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. On réalise 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) (?).

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 (?).

La bande de fréquence \mathcal{B} de $L << n$ fréquences fondamentales contiguës centrées en ω_j , s'exprime de la manière suivante :

$$\mathcal{B} = \left\{ \omega^* : \omega_j - \frac{m}{n} \leq \omega^* \leq \omega_j + \frac{m}{n} \right\}$$

avec $L = 2m + 1$ impair, choisi de manière à ce que les valeurs spectrales dans l'intervalle \mathcal{B} soient approximativement égales à la densité spectrale $f(\omega)$. Les valeurs du spectre à l'intérieur de cette bande doivent être relativement constante pour que le spectre lissé soit un bon estimateur.

Le périodogramme lissé, ou périodogramme moyen, est défini comme la moyenne des valeurs du périodogramme sur la bande \mathcal{B} .

$$\bar{f}(\omega) = \frac{1}{L} \sum_{k=-m}^m I(\omega_j + k/n)$$

Pour de grande valeurs de n , les périodogrammes moyens sont distribués comme des variables aléatoires indépendantes $f(\omega)\chi_2^2/2$ pour $0 < \omega < 1/2$, ainsi

$$\frac{2L\bar{f}(\omega)}{f(\omega)} \sim \chi_{2L}^2$$

On parle alors de fenêtre pour faire référence à la largeur de l'intervalle de fréquence définit par $B = L/n$.

La FFT est redondante lorsque n est un multiple de plusieurs facteurs de 2,3 ou 5. Pour profiter de cette propriété, il est commun de rajouter des 0 pour atteindre l'entier composite le plus proche n' . Dans ce cas, il faut remplacer $2L$ par $2L/n'$. On définit les degrés de liberté ajustés :

$$df = \frac{2Ln}{n'}$$

On peut ainsi obtenir une approximation d'un intervalle de confiance à $100(1 - \alpha)\%$ de la forme :

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

Il est commun d'observer des pics mineurs au niveau des harmoniques car la plupart des signaux ne

sont pas des sinusoïdes 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, on considère que la borne inférieure de l'intervalle de confiance du pic d'intérêt doit être supérieure au bruit pour que ce pic soit considéré comme significatif.

Lors de la création de 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. On choisit d'utiliser une cloche cosinus de la forme :

$$h_t = 0.5 \left[1 + \cos\left(\frac{2\pi(t - \bar{t})}{n}\right) \right]$$

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. On peut modéliser cette évolution par des fonctions. Si une fonction peut être exprimée par des coefficients, on cherche alors à trouver 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 passer des données d'échantillonnages qui sont discrètes à des fonctions continues sur l'intervalle que l'on souhaite comparer. Dans le cas des données HYDRO et PICAONANO on s'intéresse à des unités annuelles, une fonction correspond alors à une année. Dans le cas des données CTD, une fonction correspond à un profil à la date d'échantillonnage.

Pour obtenir une fonction $x(t)$, on utilise un système de fonctions de bases. C'est un ensemble de K fonctions ϕ_k qui sont connues et indépendantes les unes des autres. Une association linéaire de ces K fonctions nous 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 étant 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. On suit une méthode subjective qui consiste à tester plusieurs valeurs de λ et à sélectionner celle avec laquelle on obtient le lissage le plus adapté. On fait le choix d'utiliser comme fonctions de bases des splines cubiques, comme suggéré par ?.

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, on va 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 composante principale et $\hat{\xi}_l(t)$ sa fonction propre associée, il est possible de les exprimer de la manière suivante (?) :

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

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 est de 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. On réalise cette transformation à l'aide d'ACP fonctionnelles présentées précédemment. On réalise une ACP 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. On donne un poids de $\sqrt{\lambda_1}/\sqrt{\lambda_1} = 1$ à la première composante et un poids de $\sqrt{\lambda_2}/\sqrt{\lambda_1}$ à la deuxième.

On a donc de nouvelles observations qui correspondent chacune à une année à un site. Pour chaque observation, on 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 variables non-attribuées, on utilise la fonction *inputePCA* du paquet *missMDA*. 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 (?). 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'observations centré et réduit, on réalise l'AFD 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 (?). On définit les variables canoniques comme étant 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 trou-

ver les combinaisons linéaires qui maximisent la corrélation entre ces variables canoniques (corrélation canonique) (?).

On utilise cette méthode pour explorer les corrélations entre les données biologiques et les données hydrologiques. Dans un premier temps, on s'intéresse aux relations entre les données d'abondance et les concentrations en nutriments. Dans un second temps, on s'intéresse à un possible lien entre les valeurs de SSC et les concentrations en nutriments.

Comme pour l'AFD, on ne va pas s'intéresser aux observations ponctuelles, mais à l'évolution d'une variable au cours d'une année. On a donc recours aux scores des ACP utilisés lors de l'AFD. Au cours des analyses, on a remarqué que souvent, la première composante principale rend compte d'un effet moyen. Si c'est en effet le cas, on récupère 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

Pour chaque groupe du phytoplancton et pour chaque station, on réalise une régression polynomiale locale sur les observations d'abondance. Après avoir essayé plusieurs fenêtres de lissage, on établit que 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 semblent présenter une saisonnalité annuelle plus ou moins marquée.

Pour investiguer la saisonnalité on réalise le périodogramme de chaque régression. Le périodogramme des *Synechococcus* à Banyuls est présenté en Figure 7. Le périodogramme échelonné est utilisé principalement pour la construction du périodogramme moyen auquel on s'intéresse surtout. Celui-ci identifie

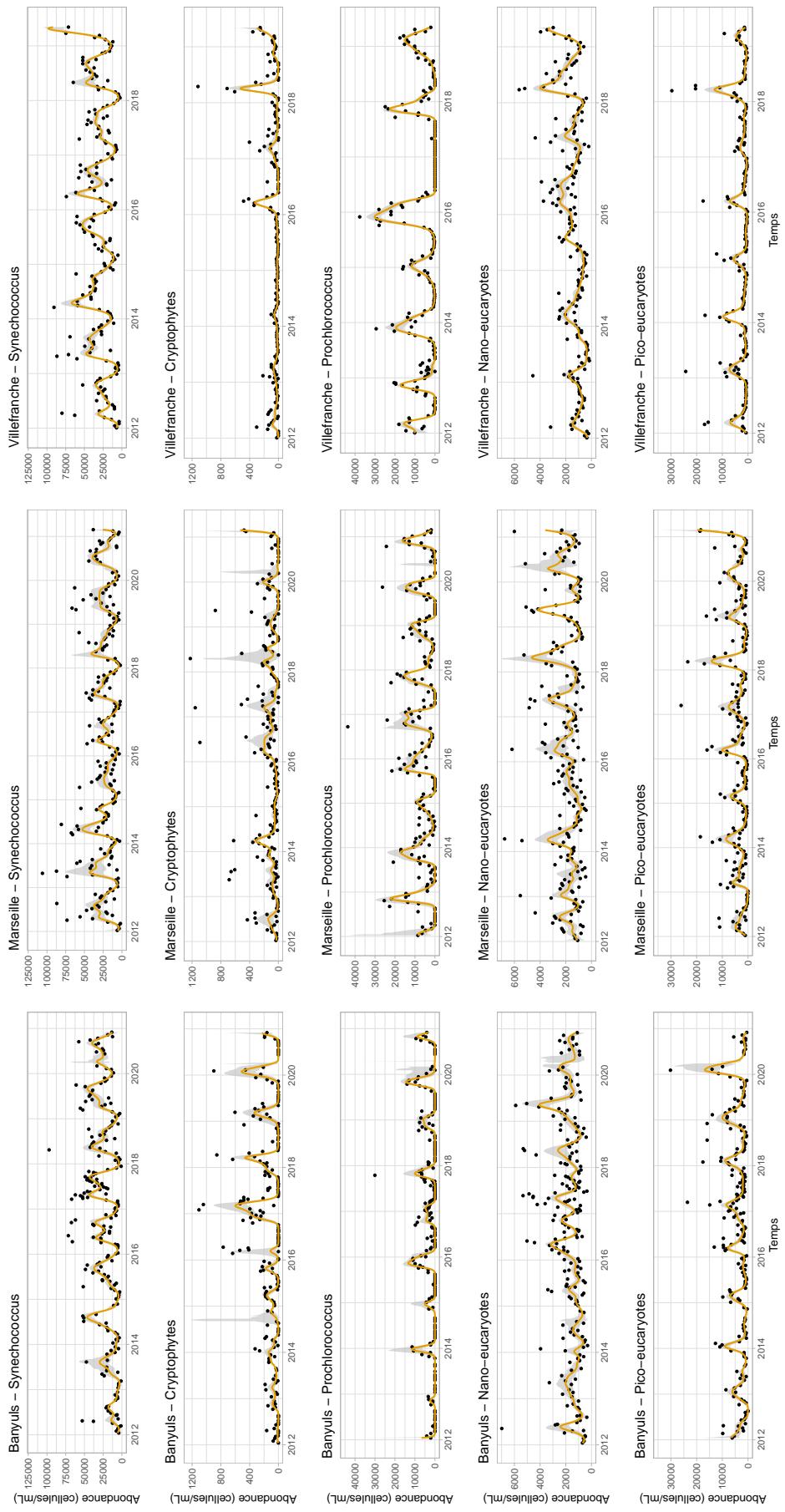


Figure 6 : Séries temporelles des abondances des groupes de phytoplancton pour les trois stations méditerranéennes. Les points noirs correspondent aux données brutes, la courbe orange à la régression quadratique locale avec noyau gaussien avec une fenêtre de 40 jours. Elle est associée à son intervalle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

clairement deux pics 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, on a également une saisonnalité significative de 6 mois.

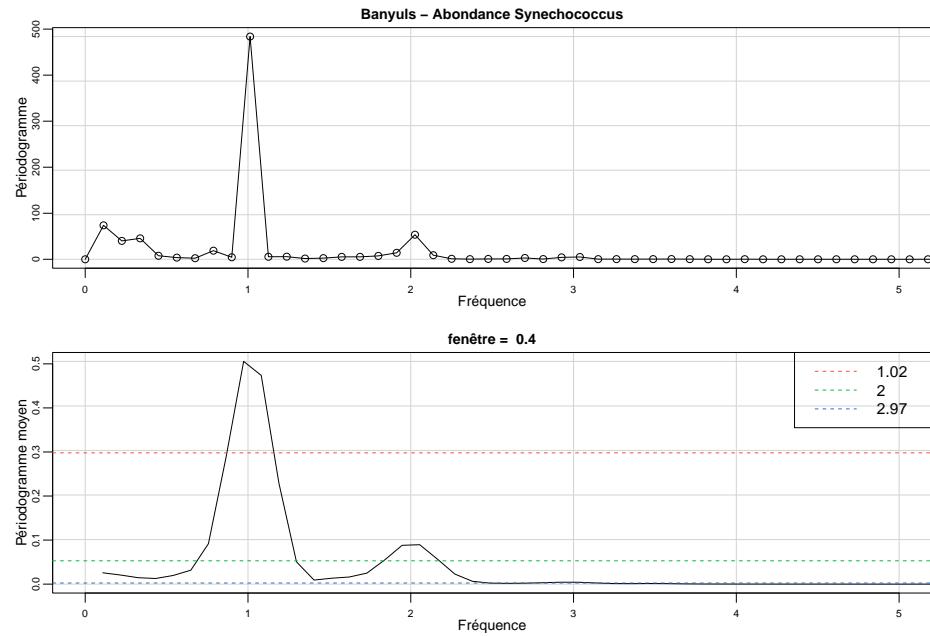


Figure 7 : Périodogramme échelonné (haut) et périodogramme moyen (bas) de la série d'abondance de *Synechococcus* à Banyuls. Les axes de fréquences sont gradués par multiple de $\Delta = 1\text{cycle}/1\text{an}$. Sur la figure du bas, les lignes pointillées correspondent à la borne inférieure de l'intervalle de confiance à 95% des pics identifiés dans la légende.

3.1.2 Température

Les données CTD des profils de température à Marseille ne sont pas observées à intervalle de profondeur régulier. On a environ une observation tous les 0,2 ou 0,3m. On réalise une interpolation pour avoir des données régulières tous les 0,5m de 0,5 à 55m de profondeur. On a donc 112 points pour chaque profils verticaux. On fait le choix de prendre deux fois moins de fonctions de bases que de points. On réalise donc un lissage avec pénalisation de la courbure avec 55 fonctions de bases. Plusieurs paramètres

de lissage λ sont testés et le lissage obtenu avec chaque paramètre est visualisé pour 8 profils choisis au hasard. On fait le choix de conserver $\lambda = 0.1$ qui produit le lissage visualisé en Figure 8.

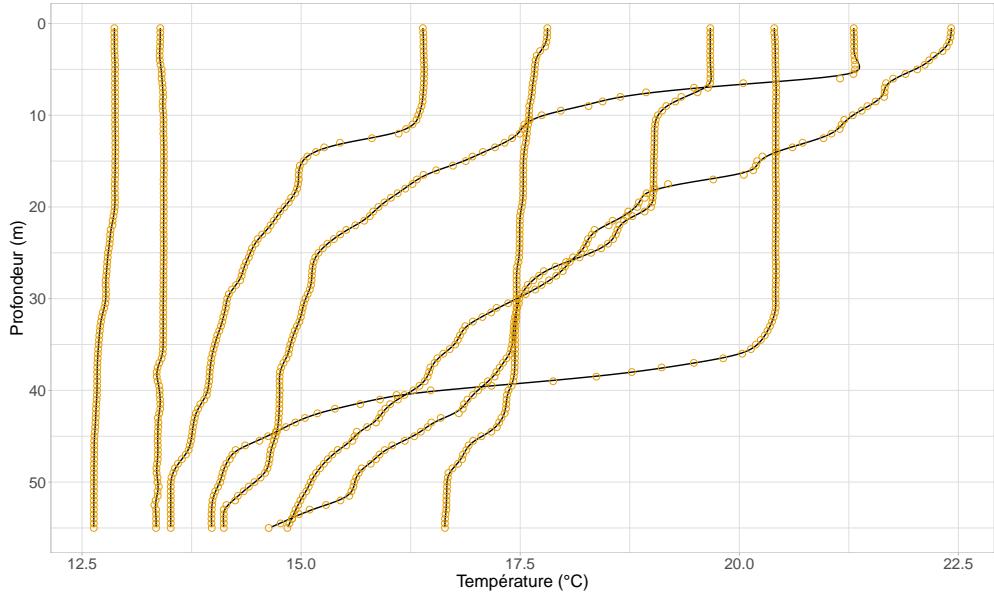


Figure 8 : Profils verticaux de température à Marseille après lissage avec pénalisation, $\lambda = 0, 1$, à partir des points oranges pour 8 dates choisies au hasard.

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 en a) de la Figure 9. 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 bidimensionnelle 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. Le b) de la Figure 9 représente sous forme de séries temporelles les scores des deux premières composantes principales pour chaque profil CTD observé. On observe une saisonnalité annuelle 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 l'on observe des valeurs négatives plus extrêmes.

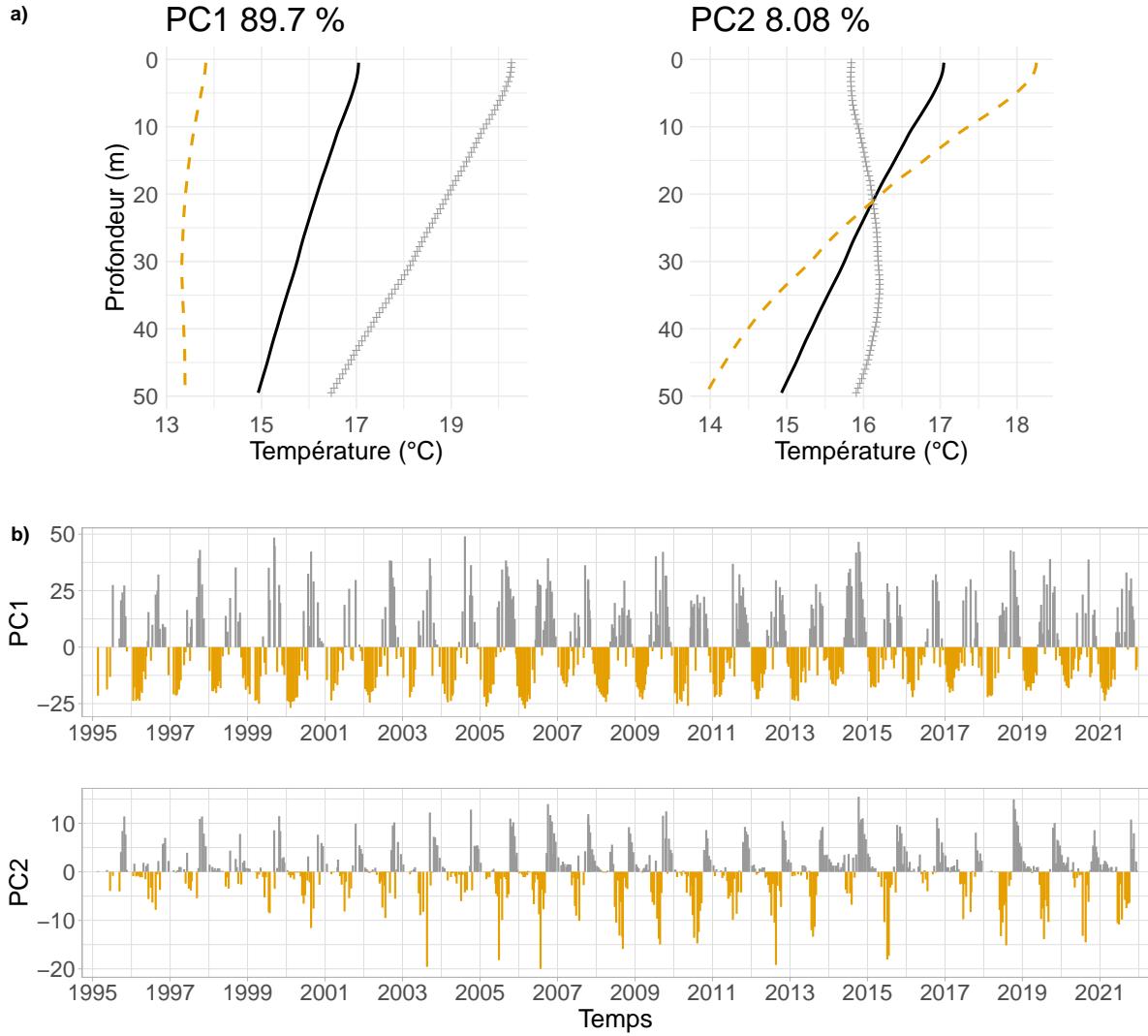


Figure 9 : a) Deux premières composantes principales de l'ACP fonctionnelle des profils de température à Marseille, exprimées comme perturbation de la moyenne. La courbe noire correspond au profil moyen. En croix grises est représenté l'effet d'une perturbation positive et en tirets orange est représenté 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. On récupère les séries temporelles des scores des PC1 et PC2 de l'ACP fonctionnelle précédente. 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, on réalise une moyenne mensuelle des scores pour chaque année. On a donc 12 observations par an. Pour plusieurs années, certains mois ne comportent aucune observation. On utilise alors le paquet *fdapace* pour pouvoir 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 noir 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 l'on a des valeurs positives du PC2. Maintenant, si l'on regarde l'effet des composantes principales de la dernière ACP, on voit que 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. En visualisant les scores de l'ACP sur deux dimensions, on voit 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

On cherche maintenant à distinguer les stations en fonction de l'évolution de chaque variable au cours de l'année. On s'intéresse ici aussi, à des unités annuelles.

3.2.1 Abondance

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

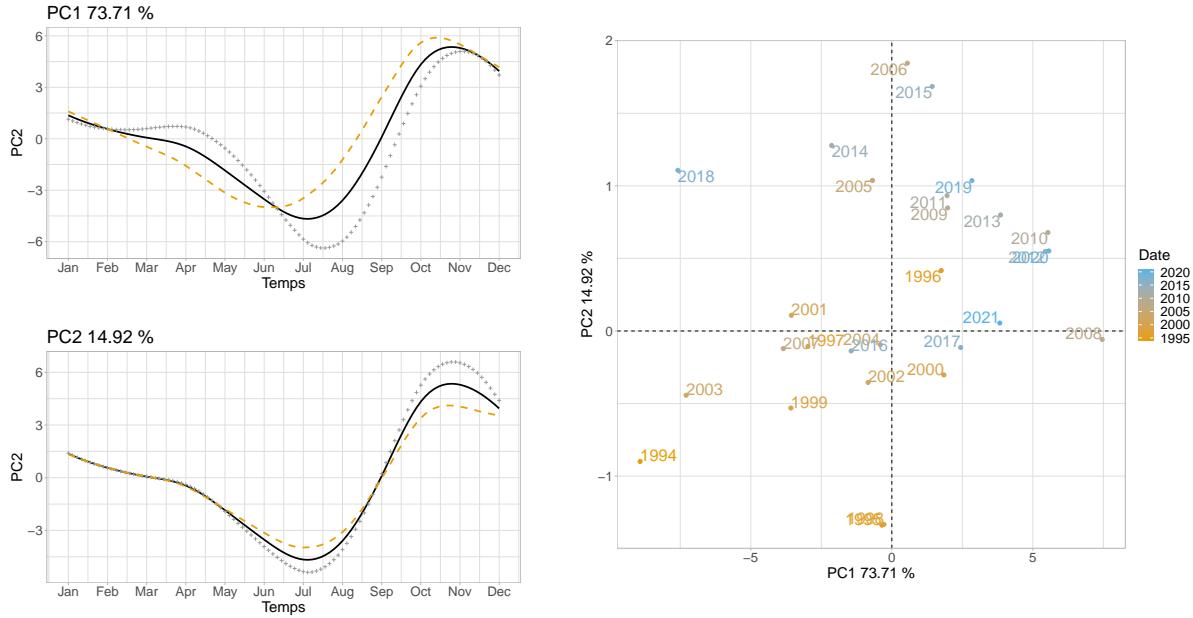


Figure 10 : (gauche) Deux premières composantes principales de l'ACP fonctionnelle de la PC2 des profils verticaux de température à Marseille, exprimées comme perturbation de la moyenne. La courbe noire correspond au profil moyen. En croix grises est représenté l'effet d'une perturbation positive et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

une diminution. L'inertie des deuxièmes composantes principales des ACP sur les abondances varie entre 8,7% à 25,8%. On choisit de 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 10 variables quantitatives qui correspondent aux scores des deux premières composantes principales des ACP pour chaque groupe du phytoplancton.

Les résultats de l'AFD sur ce tableau est présenté en Figure 11. 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 Villefranche. 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. Le PC1 des *Prochlorococcus* est celui 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 contribuent 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 pic plus importante de janvier

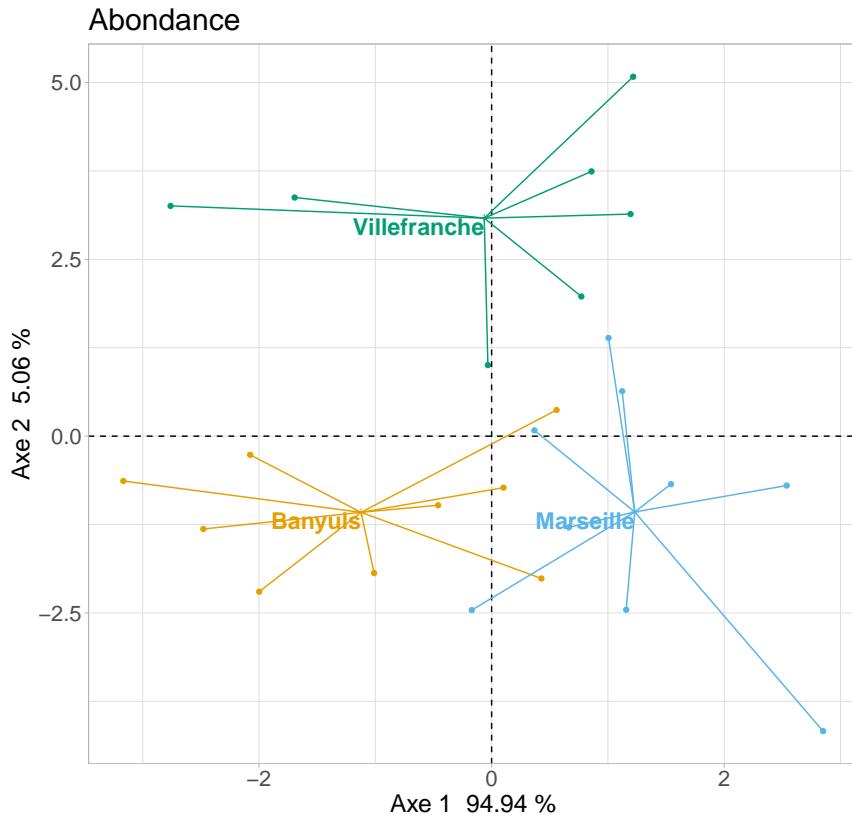


Figure 11 : Graphique des individus de l'AFD des unités annuelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.

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

Comme pour l'abondance, on réalise une ACP 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 est majoritairement inférieure à 10%. On choisit 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

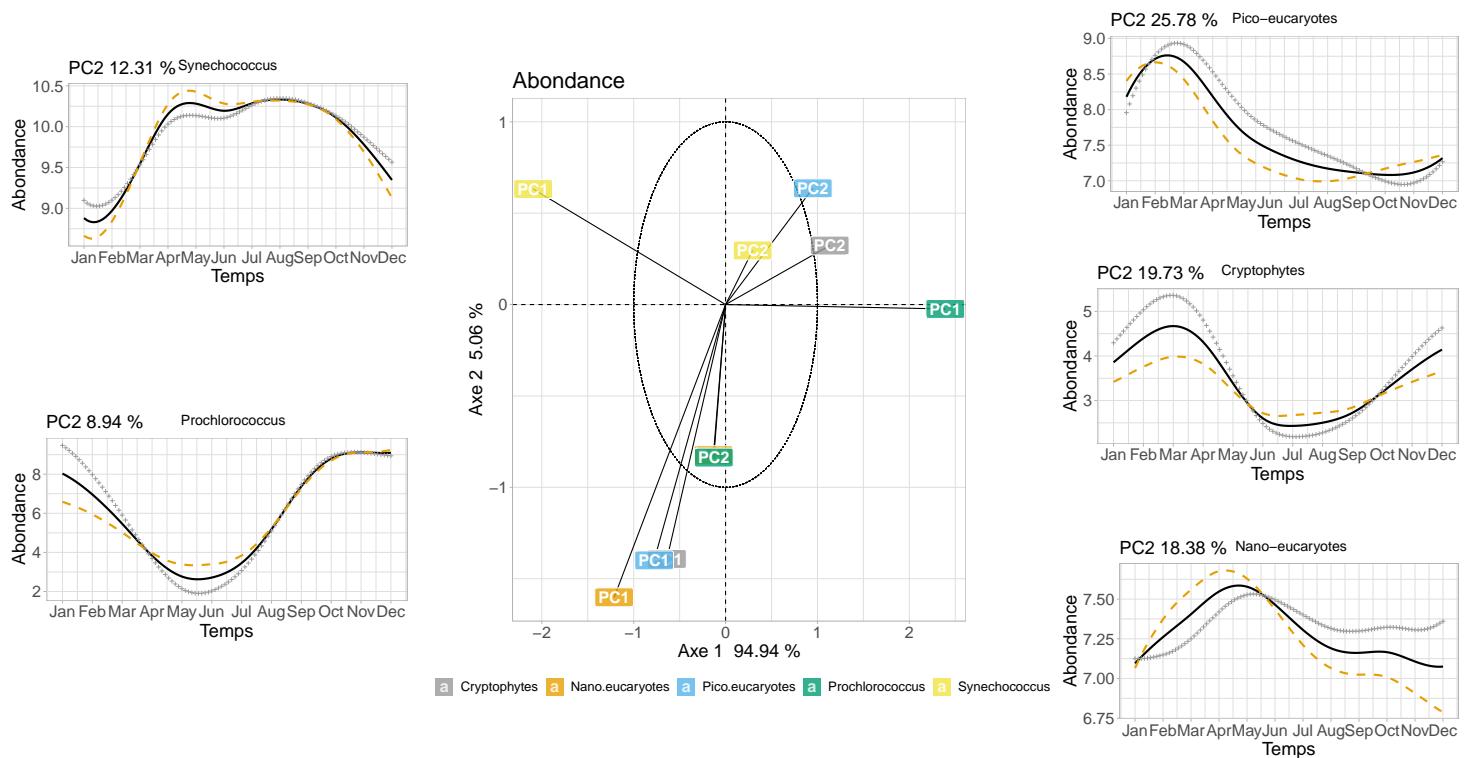


Figure 12 : (centre) Graphique des corrélations des variables de l'AFD des unités annuelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes. (côtés) Deuxièmes composantes principales des ACP fonctionnelles d'abondance du phytoplancton exprimées comme perturbation 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.

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. Si on regarde la corrélation des variables, on voit qu'une forte SSC des nano-eucaryotes s'oppose à une forte SSC des *Synechococcus* et des pico-eucaryotes. Le PC1 des *Prochlorococcus* est celui qui semble contribuer le plus à l'axe 2. À Banyuls on aurait donc des nano-eucaryotes avec 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).

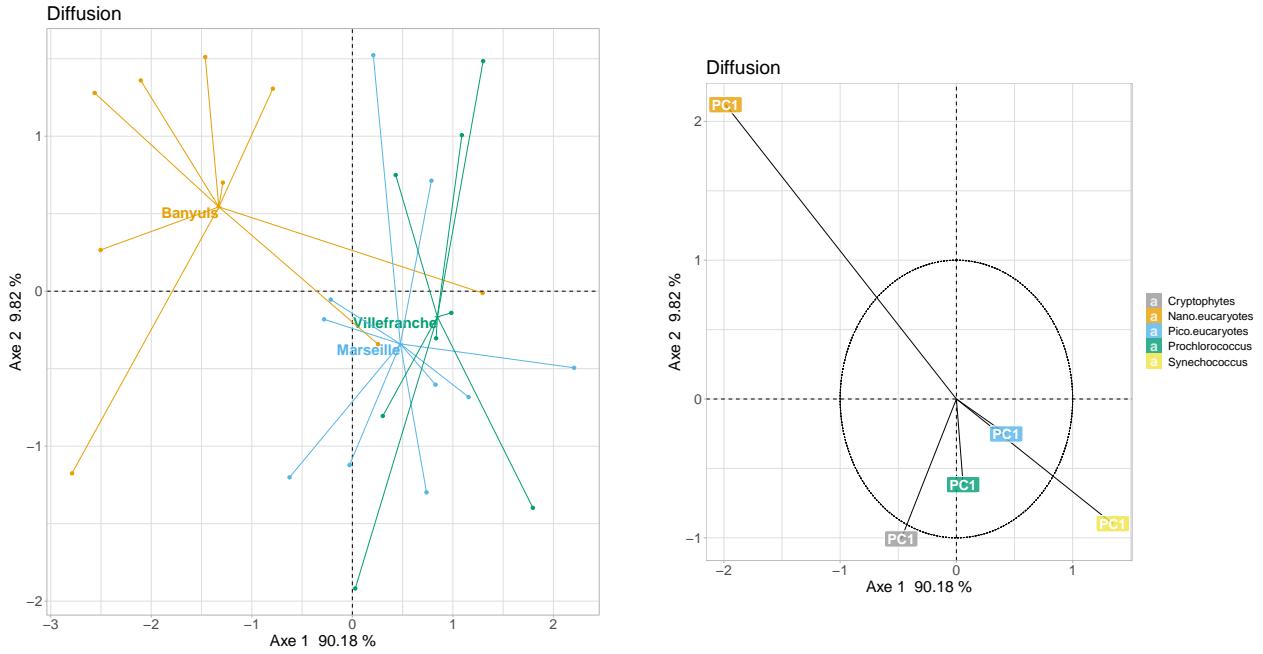


Figure 13 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l’AFD des unités annuelles de diffusion lumineuse des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.

3.2.3 Nutriments

Cette fois, au lieu d’avoir 5 groupes de phytoplancton, on s’intéresse à 5 nutriments différents (NH_4 , NO_3 , NO_2 , PO_4 , Si(OH)_4). On réalise une ACP pour chaque nutriment toutes stations confondues, avec les 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%. On choisit 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 nutriment.

Les résultats de l’AFD sur ce tableau est présenté en Figure 14. 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. Si on regarde la corrélation de variables, on voit que de fortes concentrations en nitrate

et en ammonium s'opposent à de fortes concentrations en phosphate et en nitrite. On aurait donc le site de Banyuls qui serait plutôt riche en nitrate et en ammonium et pauvre en phosphate et nitrite, à l'inverse de Villefranche. Le site de Marseille serait particulièrement riche en silice, par rapport aux deux autres, mais avec des caractéristiques plus proches de Banyuls.

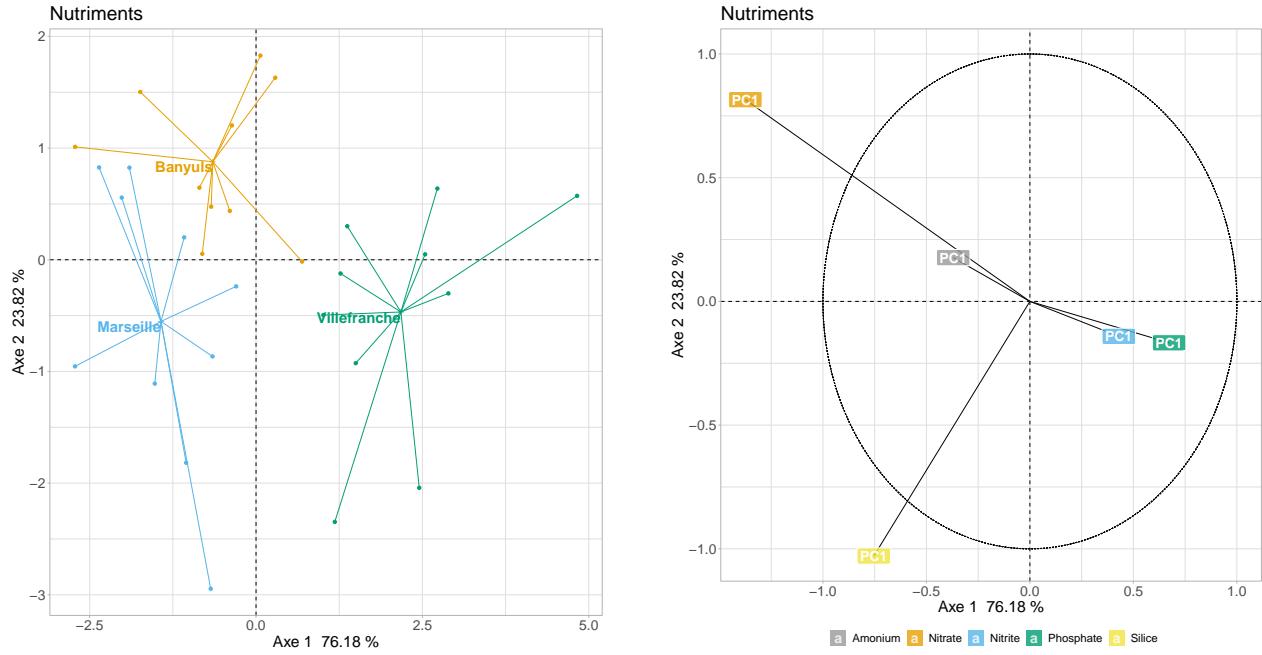


Figure 14 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'AFD des unités annuelles de concentration des nutriments pour les trois stations SOMLIT méditerranéennes.

3.3 Co-variations

3.3.1 Abondance et Nutriments

On récupère les scores de premières composantes principales des ACP sur les données d'abondance du phytoplancton et de concentration des nutriments. Elles rendent toutes compte d'un effet moyen (Annexes A et B). On a 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 15. 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 15 seules

les deux premières dimensions sont représentées. Le graphique de droite permet de comprendre les relations entre les variables. Plus elles sont proches et 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. On observe une opposition entre les sites de Villefranche et Banyuls, avec d'un côté un site avec une concentration plus importante en PO₄ et une abondance plus importante des *Prochlorococcus* et de l'autre, un site avec une concentration plus élevée en NO₃ et une abondance des pico-eucaryotes plus importante.

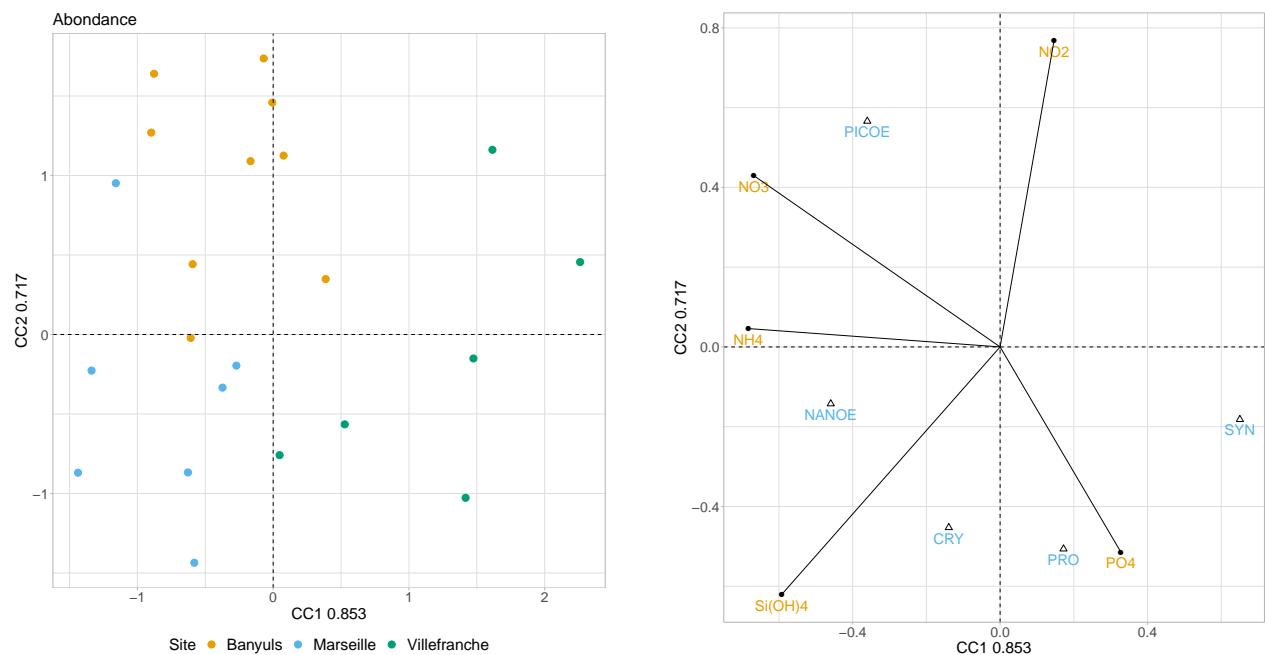


Figure 15 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l'ACC des unités annuelles de concentration des nutriments et de l'abondance des groupes du phytoplancton pour les trois stations SOMLIT méditerranéennes. Le premier axe correspond à la dimension 1 avec 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

On récupère les scores de 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). On a donc deux jeux de données, d'une part les données de 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 16. 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 16 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. On remarque 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. En explorant les dimensions suivantes, on note qu'elles ne sont pas mieux représentées. La SSC des *Synechococcus* ne semble donc corrélée à aucun des nutriments.

4 DISCUSSION

text

5 CONCLUSION

text

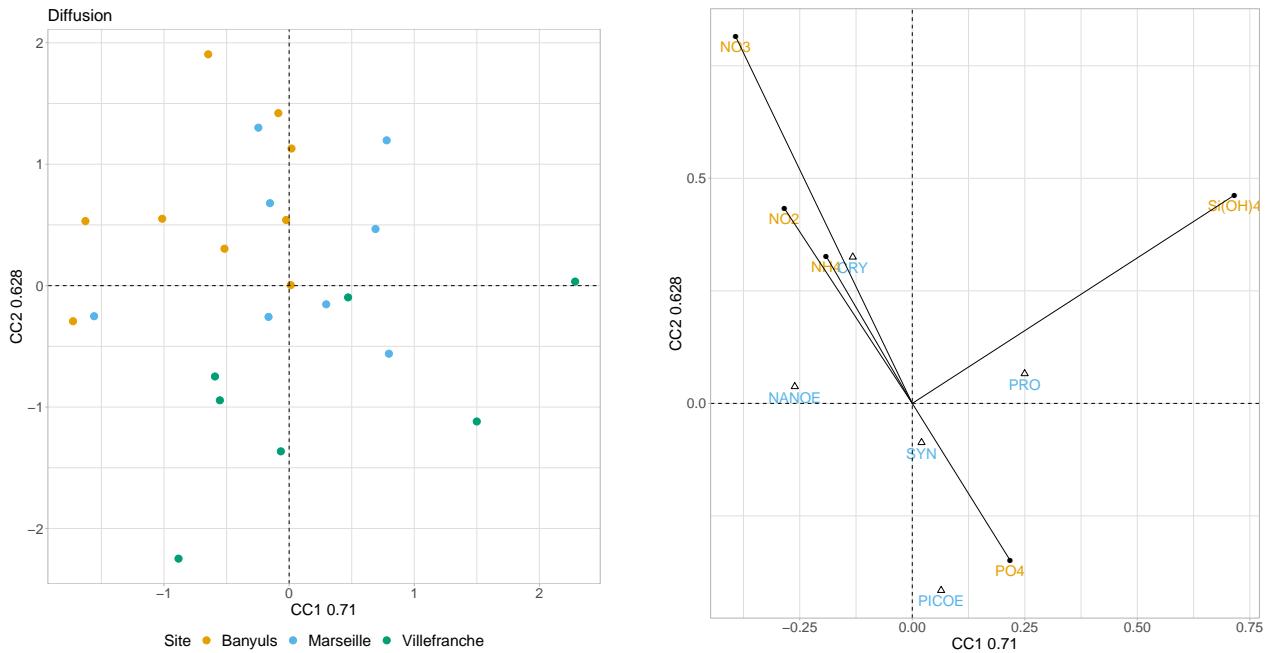


Figure 16 : (gauche) Graphique des individus et (droite) graphique des corrélations des variables de l’ACC des unités annuelles de concentration des nutriments et de la diffusion lumineuse des groupes du phytoplancton pour les trois stations SOMLIT méditerranéennes. Le premier axe correspond à la dimension 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 (NA-NOE)) par des triangles.

LISTE DES FIGURES

- | | | |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| 1 | Carte des stations SOMLIT méditerranéennes : Sola à Banyuls, Frioul à Marseille et Point B à Villefranche. | 4 |
| 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 <i>Prochlorococcus</i> (PRO), en jaune les <i>Synechococcus</i> (SYN) en bleu clair les pico-eucaryotes (PICOE), en gris les Cryptophytes (CRY) et en orange les nano-eucaryotes (NA-NOE). Modifié depuis ? | 7 |

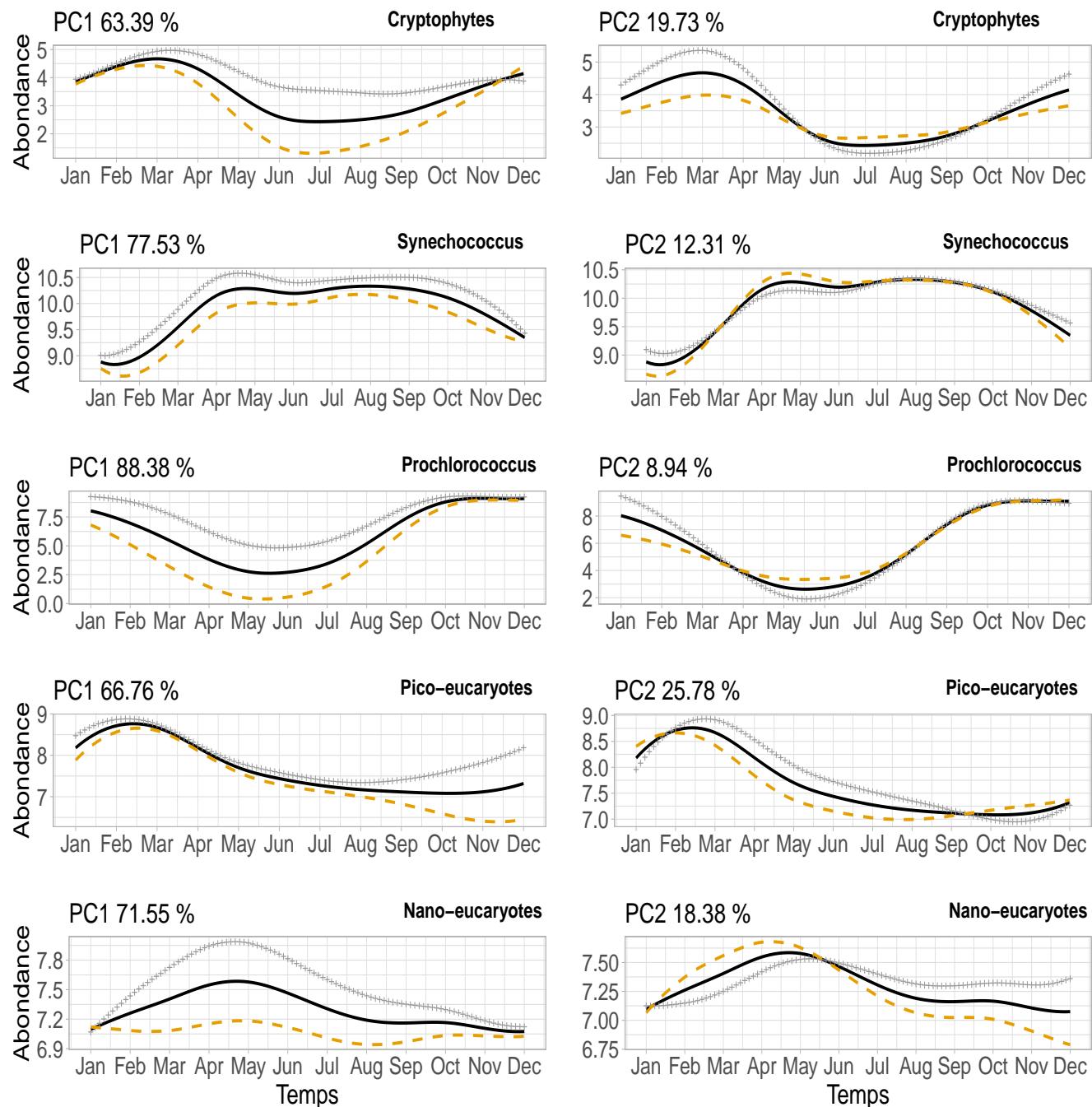
3	Séries temporelles des données hydrologiques de surface pour les trois stations SOMLIT méditerranéennes confondues. La ligne pointillé bleue correspond à la mise en place du protocole national pour la variable présentée. La ligne pointillée orange correspond au début du jeu de données PICONANO. Les éventuelles données aberrantes sont représentées en rouge.	8
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.	9
5	Auto-fluorescence rouge moyenne <i>versus</i> diffusion lumineuse à 90° moyenne de chaque groupe de phytoplancton pour chaque prélèvement pour les trois stations SOMLIT méditerranéennes.	9
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. .	17
7	Périodogramme échelonné (haut) et périodogramme moyen (bas) de la série d'abondance de <i>Synechococcus</i> à Banyuls. Les axes de fréquences sont gradués par multiple de $\Delta = 1\text{cycle}/1\text{an}$. Sur la figure du bas, les lignes pointillées correspondent à la borne inférieure de l'intervalle de confiance à 95% des pics identifiés dans la légende.	18
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.	19
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.	20

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.	22
11	Graphique des individus de l'AFD des unités annuelles d'abondance des groupes de phytoplancton pour les trois stations SOMLIT méditerranéennes.	23
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.	24
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.	25
14	(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.	26
15	(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 (<i>Prochlorococcus</i> (PRO), <i>Synechococcus</i> (SYN), pico-eucaryotes (PICOE), Cryptophytes (CRY) et nano-eucaryotes (NANO)) par des triangles. .	27

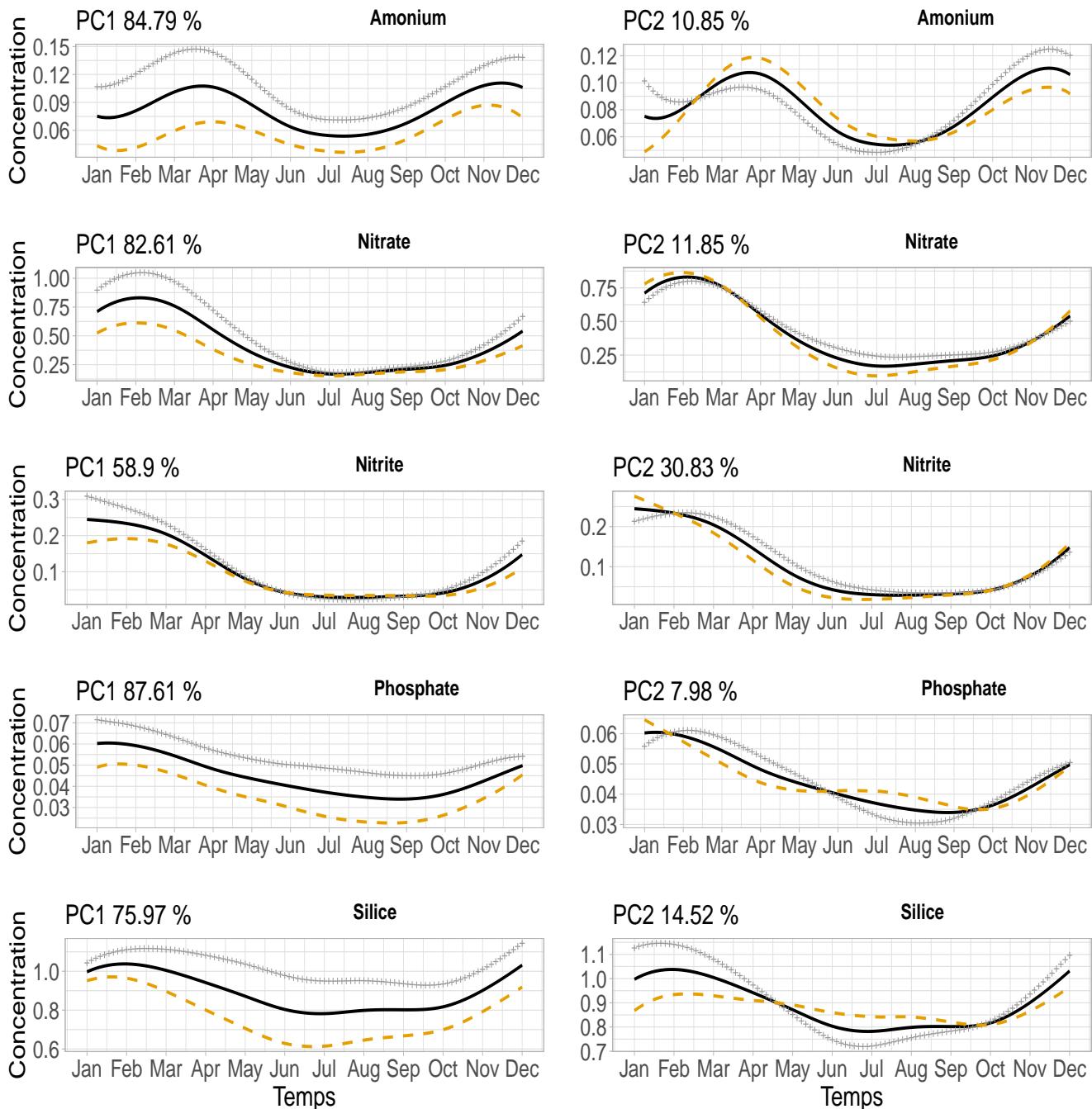
LISTE DES TABLEAUX

ANNEXES

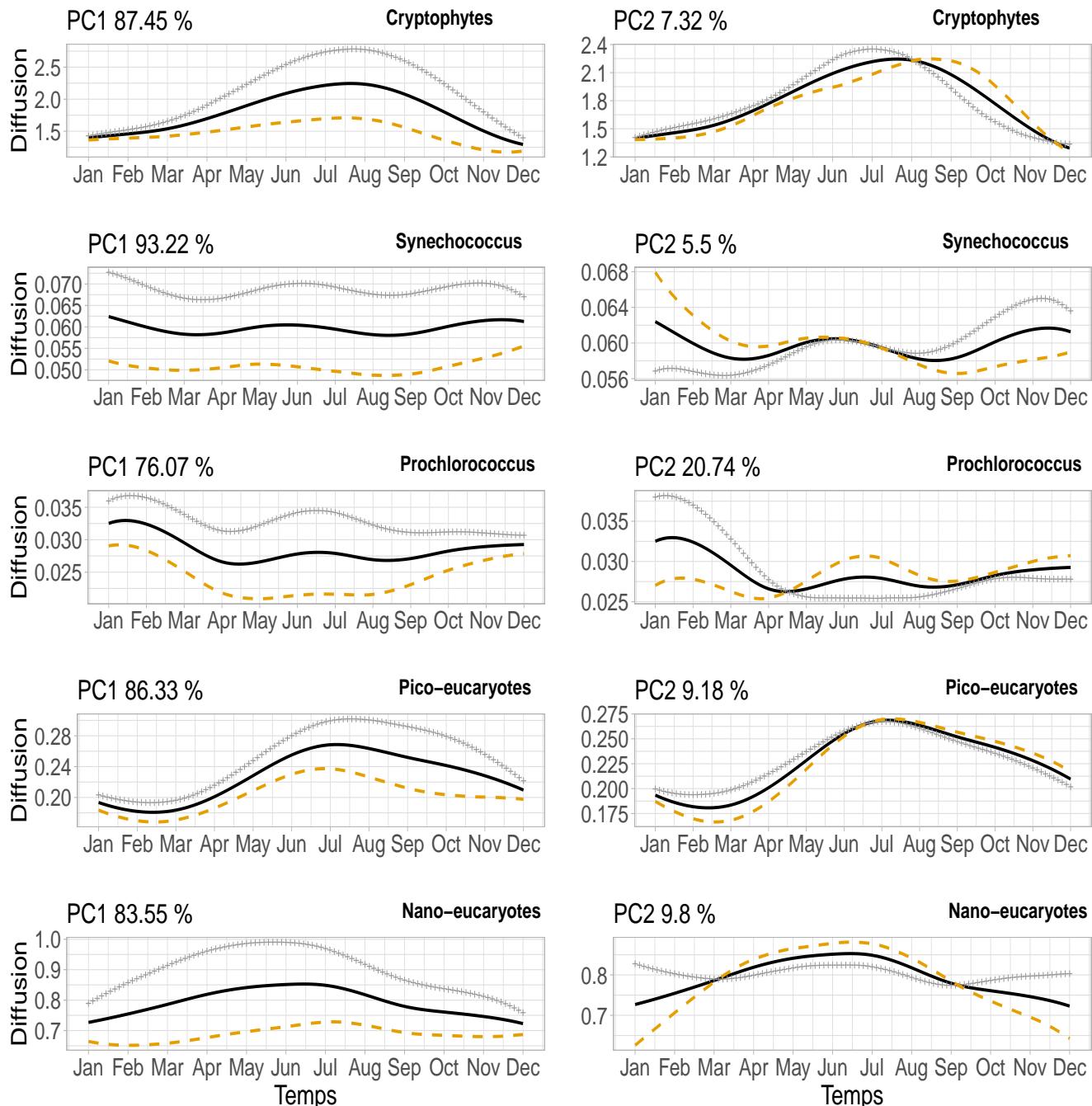
A. Deux premières composantes principales des ACP fonctionnelles des données d'abondance pour chaque groupe de phytoplancton des trois stations SOMLIT méditerranéennes, exprimées comme perturbation de la moyenne. Les courbes noires correspondent aux courbes moyennes. En croix grises est représenté l'effet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



RÉSUMÉ