



INSTITUT UNIVERSITAIRE DE TECHNOLOGIE

UNIVERSITÉ DE BRETAGNE SUD

Licence Statistique et Informatique pour la Santé

Département Statistique et Traitement Informatique des Données



ÉTUDE DES EFFLORESCENCES DE *PSEUDO-NITZSCHIA*
EN RELATION AVEC LES PARAMÈTRES ENVIRONNEMENTAUX
DANS LE MOR BRAZ



P.nitzschia australis

Source : Jacky Chauvin,
LER-MPL La Trinité-sur-Mer, IFREMER

IFREMER

Laboratoire Environnement et Ressources Morbihan-Pays de la Loire

Station de la Trinité-sur-Mer

Encadrants : Elodie FLEURY, Jean-Pierre ALLENOU et Nathalie COCHENNEC-LAUREAU



Auréliе FONTAINE

Année universitaire 2012-2013

Remerciements

Elodie Fleury, Jean-Pierre Allenou, Jacky Chauvin et Natalie Cochennec-Laureau pour leurs conseils et leur apport de connaissances sur le phytoplancton et *Pseudo-nitzschia* ;
Jean-Yves Stanisière pour ses bonnes idées et sa réflexion sur le sujet ;
Michaël Retho pour son aide sur les données physico-chimiques ;
Florian Enez, Tania Hernandez-Fariñas, Anne Thébault, Bérangère Husson pour leurs conseils et leur aide en statistique ;
Jean-François Bouget et Aimé Langlade pour les sorties en mer ;
Elizabeth Abily et Guylaine Le Mouroux pour leurs sourires chaque matin ;
Toute l'équipe de la station Ifremer de la Trinité-sur-Mer pour leur bonne humeur et leur gentillesse dont ils m'ont fait part tout au long de ce stage.

Sommaire

I.	INTRODUCTION.....	5
II.	LA STRUCTURE D'ACCUEIL.....	7
III.	MATÉRIEL ET MÉTHODES	8
III.1.	Le site d'étude	8
III.2.	Les données du RÉPHY.....	9
III.2.1.	Mesure de l'abondance du phytoplancton	9
III.2.2.	Mesure des variables environnementales	10
III.2.3.	La gestion de ces données via Quadrige ²	10
III.3.	Les données hors RÉPHY	11
III.3.1.	Données satellite	11
III.3.2.	Données météorologiques	11
III.3.3.	Données hydrologiques	11
III.3.4.	Données simulées.....	12
III.4.	Les méthodes statistiques choisies	12
III.4.1.	Le GAM.....	12
III.4.2.	Analyse descriptive de l'abondance de Pn.....	13
III.4.3.	Analyse bivariée.....	14
III.4.4.	Analyse multivariée	15
III.5.	Logiciels utilisés	17
IV.	RÉSULTATS	18
IV.1.	Réalisation de la base de données exploitable	18
IV.1.1.	Mises en forme des données brutes.....	18
IV.1.2.	Les variables calculées et ajoutées à la base.....	21
IV.2.	Description de l'abondance de Pn.....	24
IV.2.1.	Visualisation graphique.....	24
IV.2.2.	Saisonnalité et tendance.....	24
IV.2.3.	Analyse des efflorescences de Pn à OL et MeR	25
IV.3.	Analyse bivariée	28
IV.3.1.	Vérification de la linéarité des logits.....	28
IV.3.2.	Type de relation entre l'abondance de Pn et les variables environnementales	28

IV.4.	Les modèles explicatifs et prédictifs.....	29
IV.4.1.	Modèle expliquant l'apparition.....	29
IV.4.2.	Modèle expliquant la croissance de Pn.....	33
IV.4.3.	Combinaison des 2 modèles.....	35
V.	Discussion et conclusion.....	38
VI.	Liste des abréviations.....	44
VII.	Table des illustrations.....	45
VII.1.	Table des tableaux.....	45
VII.2.	Table des figures.....	45
VIII.	Bibliographie.....	48
IX.	Annexe.....	51
IX.1.	Annexe 1 : L'extraction des données du RÉPHY.....	52
IX.2.	Annexe 2 : Détails sur les données brutes.....	53
IX.3.	Annexe 3 : Le GAM et les méthodes de lissage.....	55
IX.4.	Annexe 4 : Détails sur la mise en forme de la base.....	56
IX.4.1.	Les taxons associés au genre P.nitzschia.....	56
IX.4.2.	Les indices écologiques.....	56
IX.5.	Annexe 5 : Boxplots de l'abondance de Pn.....	58
IX.6.	Annexe 6 : Matrice des corrélations de Spearman.....	59
IX.7.	Annexe 7 : La non-linéarité des logits.....	60
IX.8.	Annexe 8 : Détails des modèles obtenus.....	61
IX.8.1.	Annexe 8.1 : Le modèle expliquant l'apparition de Pn.....	61
IX.8.2.	Annexe 8.2 : Le modèle expliquant la croissance de Pn.....	63
IX.9.	Annexe 9 : Analyses complémentaires.....	64
IX.9.1.	Annexe 9.1 : Une autre approche pour la modélisation multivariée.....	64
IX.9.2.	Annexe 9.2 : Analyse de la flore phytoplanctonique.....	74
IX.9.3.	Annexe 9.3 : Les conditions environnementales rencontrées à OL et MeR.....	77

I. INTRODUCTION

Le **phytoplancton** est l'ensemble des micro-algues unicellulaires présentes dans l'eau de mer. Certaines espèces de phytoplancton sont toxiques. Elles produisent des **phycotoxines** qui peuvent être transférées par ingestion et/ou prédation à un autre organisme : par exemple des bivalves aux poissons, aux oiseaux, aux mammifères marins...et aux hommes. En France, les premières **intoxications humaines** chez les consommateurs de coquillages mises en relation avec du phytoplancton toxique datent de **1983** et **1984**.

L'**IFREMER (Institut Français de Recherche pour l'Exploitation de la MER)** a mis en place depuis cette date un réseau nommé **RÉPHY : Réseau national de surveillance du PHYtoplancton** et des **phycotoxines** dans les **coquillages**. Ce réseau acquiert de nombreuses données sur l'ensemble du littoral français dans l'eau et les coquillages, afin de répondre à trois objectifs :

- observer l'**ensemble des espèces phytoplanctoniques** des eaux côtières
- recenser les **efflorescences exceptionnelles**
- surveiller particulièrement les **espèces toxiques** pour les consommateurs et la présence de **toxines** dans les coquillages.

Les espèces de micro-algues toxiques, dangereuses pour l'homme et particulièrement surveillées par le RÉPHY appartiennent au genre *Dinophysis*, *Alexandrium* et *Pseudo-nitzschia* (abrégié **Pn**).

Le phytoplancton du genre Pn produit de l'**acide domoïque**, une neurotoxine provoquant le syndrome **ASP (Amnesic Shellfish Poisoning)**, aussi appelé en français IAFM (Intoxication Amnésiante par Fruits de Mer). Ce type de toxine peut engendrer des troubles **digestifs, neurologiques** et des syndromes **amnésiques**. Les symptômes peuvent être immédiats ou plus tardifs (de 2 à 48 heures après ingestion de mollusques contaminés) se traduisant généralement par des nausées, vomissements, diarrhées, pertes de mémoire, dommages cérébraux, convulsions, coma et voire mort. Les sujets les plus sensibles sont les enfants et les personnes âgées.

L'abondance de Pn évolue au cours du temps. La croissance de Pn est brutale en **printemps/été**, provoquant des **pics d'abondance**, appelés **efflorescences** ou encore **blooms**. Ces pics d'abondance sont de plus en plus **fréquents** et **intenses** sur le littoral français. Lorsque l'**abondance de Pn** est **forte** (au-delà de cent à trois cent milles cellules par litre d'eau de mer), une **alerte** est lancée et le taux d'acide domoïque dans les coquillages est alors surveillé. La directive européenne 2002/226/CE a mis en place des valeurs seuils de tolérance de la présence des toxines pour la consommation des coquillages. Lorsque la concentration en acide domoïque dépasse les 20 µg/g dans la chair des coquillages, les laboratoires de l'Ifremer en informent les autorités

déconcentrées de l'état (Préfecture) et des **arrêtés d'interdiction de pêche** sont diffusés. Cela permet de limiter le nombre de cas d'intoxications par ingestion de coquillages contaminés. Il en résulte des fermetures répétées de la pêche de coquilles Saint Jacques, particulièrement contaminées. Celle-ci a ainsi été fermée de 2009 à 2011 dans le **Mor Braz**, en Bretagne, provoquant de grosses pertes financières pour les filières de pêche. Ce sont donc des enjeux à la fois **économiques, écologiques** et de **santé publique** qui poussent à étudier ce phytoplancton.

Ce phytoplancton est cosmopolite (Hasle *et al.*, 2002 ; Fryxell *et al.*, 2003; Casteleyn *et al.*, 2008) et ses relations avec les conditions environnementales ont été beaucoup étudiées dans le monde. On sait notamment que sa croissance est influencée par la température, le rayonnement et les nutriments (Downes-Tettmar *et al.*, 2013). Toutefois, de nombreux résultats diffèrent selon les zones géographiques ciblées et ces **relations** sont donc **mal établies** (Thessen *et al.*, 2009 ; Lelong *et al.*, 2012). Il est donc important d'étudier ce phénomène et ces relations pour comprendre son origine dans le Mor Braz.

Ce stage s'inscrit dans un **projet de recherche** en cours de développement nommé **PSEUDOCCOQ**. L'objet de ce projet est d'étudier l'impact des changements environnementaux et anthropiques sur la dynamique spatio-temporelle de Pn et les mécanismes de contamination et de décontamination de la coquille Saint Jacques *Pecten maximus*. Ce projet de recherche comprend plusieurs axes :

- Axe 1 : **État des connaissances et étude des efflorescences et de l'abondance de *Pseudo-nitzschia* en relation avec les paramètres environnementaux dans le Mor Braz**
- Axe 2 : Étude de la production et de l'accumulation de l'acide domoïque
- Axe 3 : Étude du devenir de l'acide domoïque dans les coquillages et de la détoxification

Mon stage s'insère dans le premier axe d'étude et a pour objectif principal de rechercher des **liens** statistiques entre l'**abondance de Pn** et les **paramètres environnementaux** dans le **Mor Braz**.

Ce rapport se compose de plusieurs parties. Après une présentation de la structure d'accueil, le **site d'étude**, les **données brutes** utilisées et les **méthodes statistiques** choisies pour traiter ces données sont exposées. Puis la **réalisation de la base de données exploitable** est présentée. Les **résultats** sont décrits, discutés et confrontés aux connaissances acquises dans la bibliographie.

II. LA STRUCTURE D'ACCUEIL

L'IFREMER a été créé en 1984. Cet établissement public est placé sous la tutelle des ministères de l'Enseignement supérieur et de la Recherche et de l'Écologie, du Développement durable et de l'Énergie. Il contribue à la connaissance des **océans** et de leurs **ressources**. Grâce à des réseaux de suivi et de surveillance du milieu marin, l'IFREMER apporte son expertise aux professionnels et au public. Les projets et missions sont axés vers la biodiversité marine, les effets du changement climatique, la prévention des pollutions et la qualité des produits de la mer.

L'IFREMER s'organise en **quatre départements** recouvrant différents domaines d'expertise :

- Ressources Biologiques et Environnement ;
- Ressources physiques et écosystèmes de fond de mer ;
- Océanographie et dynamique des écosystèmes ;
- Infrastructures marines et numériques.

Le département **Océanographie et dynamique des écosystèmes** cible principalement la surveillance et l'observation du milieu marin, ainsi que la dynamique des écosystèmes côtiers et de l'océan à moyenne et grande échelle. **Trois unités de recherche** sont rattachées à ce département : Environnement Ressources, Dynamique de l'Environnement Côtier et Océanographie Physique et Spatiale.

L'unité de recherche **Environnement Ressources** comporte **huit laboratoires** dont celui du **Morbihan-Pays de la Loire**. Ce dernier est implanté à Nantes et à **la Trinité-sur-Mer**, où s'effectue le stage. Les missions de ce dernier portent sur :

- **l'observation, la surveillance** et le **diagnostic** de la **qualité des littoraux** ;
- **l'étude des écosystèmes littoraux et conchylicoles** ;
- les recherches régionales intégrées sur l'interaction entre le bassin versant et le littoral ;
- le suivi de la **qualité des produits de l'aquaculture** dans le milieu naturel ;
- l'expertise et **l'émission d'avis** à l'intention des services déconcentrés de l'État ;
- la valorisation et le transfert des connaissances.

Pour cela, le laboratoire de la Trinité-sur-Mer met en œuvre de nombreux **réseaux de surveillance**, notamment le RÉPHY dans la zone du Mor Braz. Ce sont donc des données acquises par le laboratoire de la Trinité-sur-Mer dans le cadre du RÉPHY qui sont traitées durant ce stage.

III. MATÉRIEL ET MÉTHODES

III.1. Le site d'étude

L'étude est réalisée sur une zone géographique précise appelée le **Mor Braz**. Celle-ci correspond à une baie située au sud de la Bretagne, délimitée par la presqu'île de Quiberon et l'estuaire de la Vilaine (Figure 1). Plusieurs **points de suivi** du RÉPHY y sont localisés.

Les points de suivi **Ouest Loscolo (OL)** et **Men er Roue (MeR)** disposent de données régulières et complètes sur le phytoplancton chaque année depuis 1984. Ces deux points sont de plus situés dans deux zones contrastées au niveau des conditions environnementales (apport en nutriments et hydrodynamisme). L'étude porte sur ces deux points de suivi.

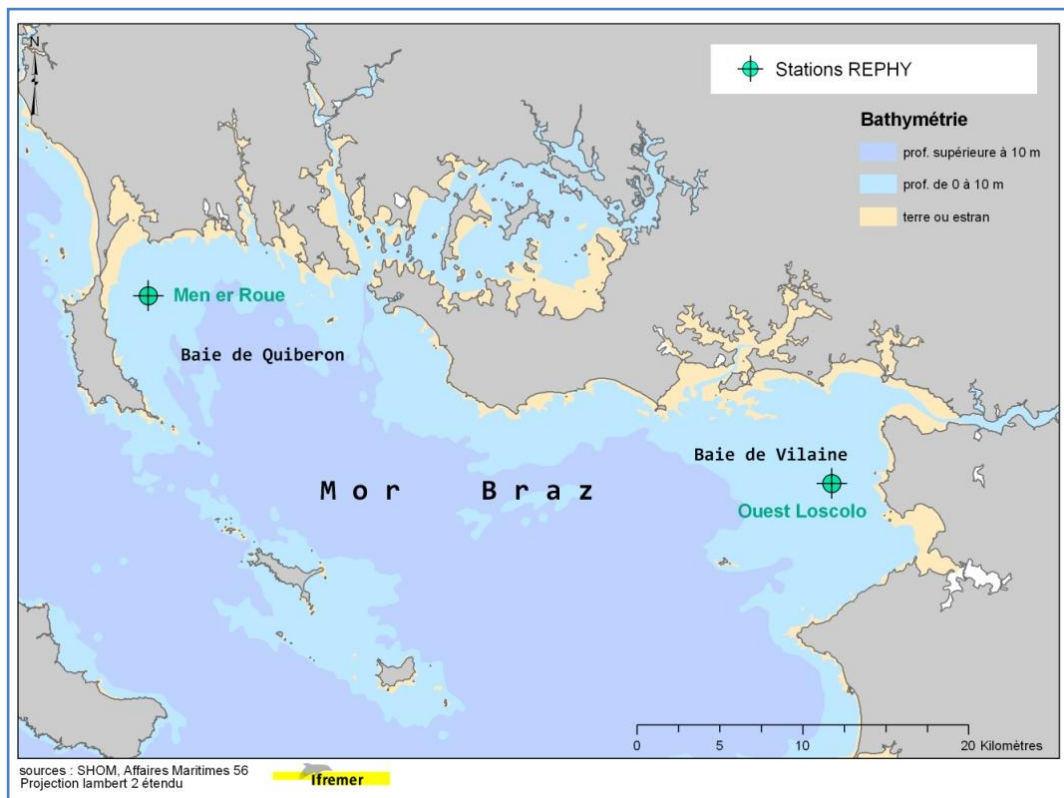


Figure 1 : Carte du Mor Braz et localisation des points de suivi du RÉPHY Ouest Loscolo et Men er Roue

III.2. Les données du RÉPHY

Les nombreuses données acquises dans le cadre du RÉPHY au niveau des points de suivi OL et MeR permettent d'étudier l'**ensemble du phytoplancton**, dont **Pn**, en apportant des informations sur leur abondance ainsi que sur leur **milieu de vie**, c'est-à-dire l'eau de mer environnante (température, salinité...).

III.2.1. Mesure de l'abondance du phytoplancton

L'acquisition des données utilisées sur le phytoplancton se fait sur des **prélèvements d'un litre d'eau de mer** réalisés au niveau des deux points de suivi (OL et MeR) à un niveau de profondeur donné. Les cellules de phytoplancton sont fixées au lugol (solution d'iodure de potassium). Au laboratoire, un technicien de l'Ifremer **identifie** et **dénombré** au microscope optique l'ensemble des cellules de phytoplancton décantées dans un **échantillon de 10 mL** de ce prélèvement (Figure 2). Le **taxon** correspondant à l'identification la plus précise de la cellule par microscope optique est noté par le technicien (genre, famille, groupe d'espèces...).

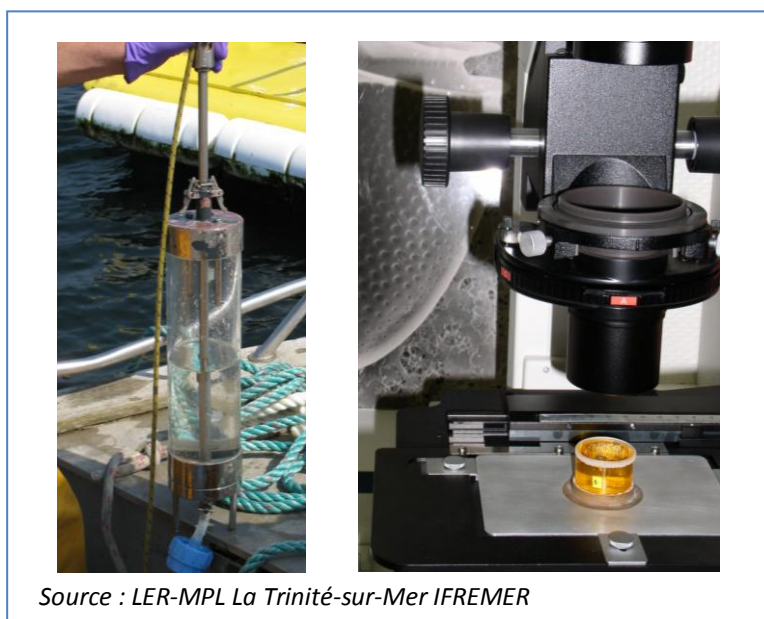


Figure 2 : Prélèvement d'eau en mer et comptage des cellules de phytoplancton

Le technicien note la **concentration en cellules de phytoplancton** par **taxon** en **cellules/litres** d'eau de mer, pour un point de suivi donné. Dans le rapport, cette variable est nommée « **Abondance** ». En fonction des objectifs recherchés, trois **types de comptage** avec des fréquences d'échantillonnage différentes sont effectués à OL et MeR :

- La **flore totale** : **toutes** les cellules phytoplanctoniques identifiables sont dénombrées. Ce type de comptage est effectué toute l'année, deux fois par mois.

- La **flore indicatrice** : seules les espèces **toxiques** et celles présentes à une abondance de plus de **100 000 cellules** par litre sont recherchées et comptées. Ceci est effectué environ deux fois par mois.
- La **flore toxique** : seules les espèces **toxiques** sont recherchées. Ce type de comptage est effectué une fois par semaine en « **périodes à risque** » et quand le seuil d’alerte fixé entre 100 000 et 300 000 cellules de phytoplancton toxique est dépassé. Lorsque ce seuil d’alerte est atteint, les toxines ASP sont dosées dans les coquillages. Les périodes à risque sont définies selon le résultat de ces dosages de toxines observé les années précédentes.

Pour chaque type de comptage, la durée entre chaque mesure peut différer et les comptages ne sont pas toujours effectués sur OL et MeR à la même date.

Pn est un phytoplancton toxique et est recherché et identifié pour chaque type de comptage (**flore totale, flore toxique et flore indicatrice**).

III.2.2. Mesure des variables environnementales

En plus du comptage du phytoplancton, l’**eau** de chaque prélèvement est aussi **analysée** (*in situ* ou en laboratoire). Cette analyse donne donc des informations sur le milieu de vie du phytoplancton étudié. Les paramètres mesurés à **chaque prélèvement du RÉPHY** sont :

- la **température** (en degré Celsius) ;
- la **salinité** (sans unité) ;
- la **turbidité** (la teneur en matière qui trouble un fluide, le rendant plus opaque, en NTU ou FNU) ;
- la **teneur en chlorophylle a** (principale forme de pigment des végétaux et un indicateur de la quantité de phytoplancton, en µg/L) abrégé par la teneur en **chlo a** ;
- la teneur en **nutriments** (composés nécessaires à la vie d’un organisme, en µg/L) :
 - **azote inorganique** (NO_2^- , NO_3^- , NH_4^+ et la somme $\text{NO}_2^- + \text{NO}_3^-$) ;
 - **phosphates** (PO_4^-) ;
 - et **silicates** (SiOH).

Un tableau renseignant les différentes méthodes de mesure et unités de mesures en fonction des périodes est disponible en Annexe 2.

III.2.3. La gestion de ces données via Quadrigé²

Les données acquises sont saisies sur **Quadrigé²**, qui est un **système d’information** développé par l’Ifremer pour gérer l’ensemble des données de surveillance du littoral. Ces données sont contrôlées par une personne autre que le saisisseur, puis validées par le correspondant local du

RÉPHY et enfin qualifiées par les gestionnaires de ce système d'information. Elles sont ensuite diffusées puis extractibles via ce système.

Toutes les données disponibles du RÉPHY sur le phytoplancton abordées dans cette partie (données quantitatives) et concernant les sites **OL** et **MeR** sont extraites de Quadrige². Un exemple d'interface d'extraction de ce système est mis en Annexe 1. La période disponible pour ces données est comprise entre **1984 et 2012**. 1652 échantillons ont été prélevés dans le cadre du RÉPHY, dont 1271 de flore totale, 20 de flore indicatrice et 361 de flore toxique. Les données hors RÉPHY

Pour compléter les données du RÉPHY, d'autres données ont été importées dans la base des données analysées. Elles sont toutes **journalières** et **spatiales** et décrivent les **conditions environnementales** rencontrées dans le **Mor Braz**.

III.3. Les données hors RÉPHY

III.3.1. Données satellite

Deux d'entre elles proviennent d'une série d'instruments d'observation scientifique couplés à un système embarqué satellitaire lancé par la NASA : **MODIS** (*Moderate-Resolution Imaging Spectroradiometer* qui peut se traduire en français par « Radiomètre spectral pour imagerie de résolution moyenne »). Elles ont une résolution spatiale de 250m à 1km et sont :

- la teneur en **matière inorganique en suspension** (en mg/m³), abrégé par la **teneur en mies** : ce paramètre donne sensiblement la même information que la turbidité ;
- la teneur en **chlorophylle a** (en mg/m³), abrégé par la **teneur en chlo a** ;
- et le **coefficient d'extinction lumineuse** (sans unité) : il indique la capacité de la lumière à pénétrer dans l'eau et plus il est élevé, moins la lumière pénètre dans l'eau (il est calculé à partir des deux paramètres précédents) ;

III.3.2. Données météorologiques

Les données importées de **rayonnement** (l'énergie lumineuse reçue, en J/cm²) autour du Golfe du Morbihan (Auray, Vannes, Sarzeau...) proviennent de la base de données de **Météo France** (mesurées au niveau de stations météorologiques de Vannes, Auray, Sarzeau...).

III.3.3. Données hydrologiques

La Loire et la Vilaine sont les fleuves les plus proches des points de suivi Ouest Loscolo et Men er Roue (Figure 1) et transportent les nutriments. Des données de **débits** de la **Loire** et de la **Vilaine** (en m³/s) au niveau des barrages respectifs de chaque fleuve ont été également importées

dans la base. Ces débits sont calculés par la **banque Hydro**, un service français d'hydrométéorologie, à partir de mesures *in situ* de hauteur d'eau et de courbes de tarage.

III.3.4. Données simulées

Le modèle hydrodynamique **Eco Mars 3D** de l'Ifremer a été créé pour prévoir les concentrations de nutriments et de production primaire. À partir notamment de données par satellites, celui-ci est capable de fournir des résultats quotidiens selon une maille carrée de quatre kilomètres de côté et sur trente niveaux entre la surface et le sol. Il a permis de simuler les dernières données importées dans la base :

- la **température** de surface (en degré Celsius) ;
- et un indicateur de **stratification** (sans unité). Plus il est élevé, plus l'eau est stratifiée, c'est-à-dire organisée en blocs/couches horizontales successives avec peu d'échanges entre chaque couche.

Toutes ces données sont quantitatives. Elles ont été importées ou simulées par plusieurs personnes travaillant à l'Ifremer. Les périodes disponibles à l'importation diffèrent pour chaque variable. Ces périodes sont renseignées en Annexe 2.

III.4. Les méthodes statistiques choisies

Au préalable, pour faciliter les traitements et la visualisation graphique, l'abondance de Pn a été **transformée en log10**. Diverses méthodes statistiques ont été utilisées pour traiter les données, chacune permettant d'atteindre un objectif précis. Le **GAM** (*Generalized Additive Model* soit en français **Modèle Additif Généralisé**) est un modèle non paramétrique choisi pour de nombreuses analyses, il est expliqué brièvement.

Les méthodes statistiques présentées ensuite concernent l'**abondance de Pn**, qui est la variable à expliquer. Elles sont **descriptives** et permettent de décrire sa **variation** dans le temps ainsi que d'**identifier** et de **décrire** les **efflorescences** de Pn dans la période d'étude. Enfin les **relations** entre l'abondance de Pn et les variables environnementales sont étudiées grâce à des méthodes d'analyse bivariée et multivariée.

III.4.1. Le GAM

Le **GAM** est un modèle qui possède des propriétés du modèle général linéaire (GLM) et du modèle additif (non paramétrique). Il vérifie :

$$E(Y) = f_1(X_1) + f_2(X_2) + \dots + f_m(X_m)$$

Y est la variable à expliquer. $E(Y)$ est son espérance (sa moyenne). X_1 à X_m sont les m variables explicatives. À chaque variable explicative/prédicteur est associée une **fonction non paramétrique**, au lieu des relations linéaires comme dans la régression multiple.

Ce modèle est **généralisé**. Cette somme de fonctions ci-dessus est « liée » à la variable d'intérêt par une fonction de lien. Par exemple, la fonction de lien en régression logistique, qui fait parti des GLM, est le logit. Finalement le GAM satisfait :

$$g(E(Y)) = f_1(X_1) + f_2(X_2) + \dots + f_m(X_m)$$

La variable Y suit une **loi spécifiée** de la famille exponentielle, c'est-à-dire résumée par une formule unique et possédant des propriétés communes (la plus connue étant la loi normale). g est la **fonction de lien**, monotone et différentiable. Ce sont deux conditions du GAM (comme pour le GLM).

L'intérêt du GAM est la possibilité d'avoir des relations **non linéaires** et difficiles à résumer par une seule équation entre la variable à expliquer et la variable explicative. Pour définir cette relation, on utilise diverses **méthodes de lissage**. Chaque fonction non paramétrique du modèle correspond à une méthode de lissage spécifiée. Pour les GAM utilisés dans ce rapport, les méthodes de lissage sont la **régression cubique pénalisée**, très utilisée pour les séries temporelles (Zuur A.F. et al., 2009), et le lissage « tensor product » pour les lissages de deux variables (interactions du modèle) (Woods S.N. et al., 2006). Des détails sont donnés en Annexe 3.

III.4.2. Analyse descriptive de l'abondance de Pn

Pour cette analyse, seules les données d'abondances de Pn, de lieu et de date sont utilisées. L'abondance de Pn varie dans le temps et ses efflorescences sont des phénomènes saisonniers. Afin d'observer cette variation temporelle non linéaire à OL et MeR, le **GAM** est utilisé comme une méthode de lissage et **outil descriptif** du comportement global de l'abondance de Pn dans le temps. La variation temporelle de l'abondance de Pn est observée plus précisément grâce un GAM expliquant l'abondance de Pn uniquement en fonction du **jour de l'année** et de l'**année**. Cette dépendance au temps a donc été analysée à deux échelles : une « **saisonnalité** » à l'échelle d'une année (la période) et une « **tendance** » à l'échelle de plusieurs années (à long terme). Ceci est fait de façon distincte pour **chaque lieu** afin de les comparer et d'avoir des séries temporelles correctes.

Afin de décrire le **nombre**, l'**intensité**, la **durée** et les **dates d'apparition** des **efflorescences de Pn**, une méthode de traitement des données a été mise au point. Elle a été créée par Tania Hernandez-Fariñas, doctorante en thèse à l'Ifremer de Nantes. Elle a été adaptée pour cette étude aux données d'abondance de Pn d'OL et MeR. Celle-ci fonctionne en plusieurs étapes.

Tout d'abord les données d'abondance de PN sont traitées distinctement sur **chaque site** (OL et MeR). Elles sont ensuite moyennées pour correspondre à des abondances de Pn par **quart de mois**.

On obtient ainsi une série temporelle presque continue. Les données manquantes sont obtenues par interpolation grâce à la régression spline, une méthode de lissage (annexe ...).

Le traitement se fait ensuite **par année**. Les abondances sont **divisées par le maximum atteint dans l'année**. Une efflorescence débute et prend fin avec le **dépassement** d'un **seuil** de 15% (choisi arbitrairement) de ce maximum. Les **dates de début** et de **fin** ainsi que la **durée** et l'**abondance maximale** pour chaque efflorescence sont conservées dans une table, ainsi que les dates de la **pente maximale** atteinte pour chaque efflorescence (où la croissance de Pn a été optimale).

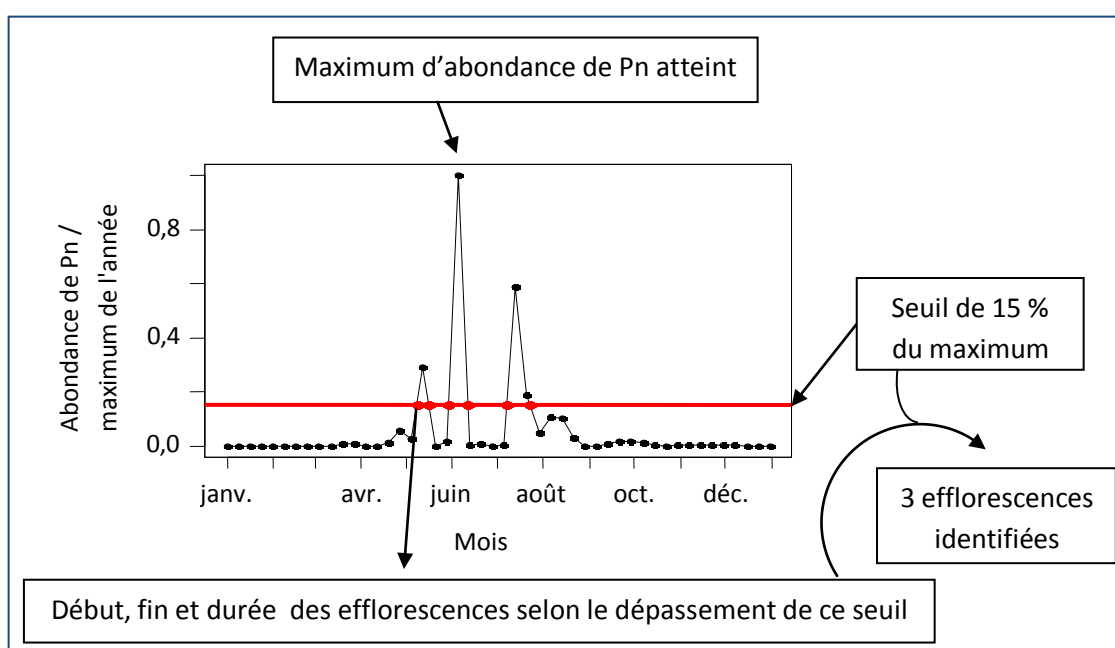


Figure 3 : Exemple de fonctionnement de l'analyse des efflorescences de Pn (données de MeR en 2006)

La Figure 3 illustre le fonctionnement de cette analyse. L'abondance de Pn est divisée par son maximum. Le maximum d'abondance prend donc la valeur de 1. Le seuil de 15% apparaît en rouge. Le premier dépassement de ce seuil par l'abondance permet d'identifier un premier bloom. Le deuxième dépassement marque sa fin. Le suivant marque un nouveau bloom, et ainsi de suite...

La table obtenue suite à cette analyse permet de **résumer** et de **décrire** les **efflorescences de Pn** sur **chaque site** et sur toute la **période d'étude**.

III.4.3. Analyse bivariée

Les **relations** entre l'abondance de Pn et chaque variable environnementale sont visualisées graphiquement. Ces relations, non linéaires, sont étudiées grâce au GAM modélisant l'abondance de Pn en fonction d'une seule variable environnementale.

L'analyse bivariée constitue une **étape préalable** à l'analyse multivariée. Elle permet d'effectuer un classement des variables selon l'intensité de leur relation avec la variable d'intérêt. Cela facilite la construction des modèles multivariés développés *a posteriori*.

III.4.4. Analyse multivariée

Des techniques de **modélisation** sont utilisées pour analyser l'**abondance de Pn en fonction des variables environnementales**. Deux approches ont été mises au point. L'approche retenue dans ce rapport décompose l'abondance de Pn en deux phénomènes à expliquer : l'**apparition** et la **croissance**/multiplication de Pn. L'autre approche intéressante ainsi que ses résultats sont exposés en annexe.

Un premier modèle est construit pour **expliquer** et prédire l'apparition de Pn, c'est-à-dire l'**absence** (abondance nulle) ou la **présence** (abondance non nulle) de Pn **en fonction des variables environnementales**. Les logits n'étant pas linéaires, un **GAM** logistique binaire est utilisé, au lieu d'une régression logistique (GLM). La loi de distribution utilisée est la loi binomiale (la variable à expliquer est binaire : 0 pour absence et 1 pour présence de Pn) et la fonction de lien utilisée est le logit (soit la fonction d'équation $y = \ln\left(\frac{x}{1-x}\right)$, où x est la probabilité de présence conditionnellement aux valeurs prises par les variables explicatives dans ce modèle).

Un deuxième modèle est utilisé pour **expliquer** et prédire la **croissance de Pn**, c'est-à-dire l'**abondance de Pn en fonction des variables environnementales** uniquement lorsque celle-ci **est présente**. Le GAM est à nouveau utilisé (relations non linéaires). La distribution spécifiée pour ce dernier GAM est la loi normale et la fonction de lien spécifiée est la fonction identité (d'équation $y=x$). Ces deux modèles sont ensuite **combinés** pour **expliquer** et prédire l'**abondance de Pn** sur l'**ensemble des données**. La Figure 4 permet de résumer la méthode d'analyse multivariée complète.

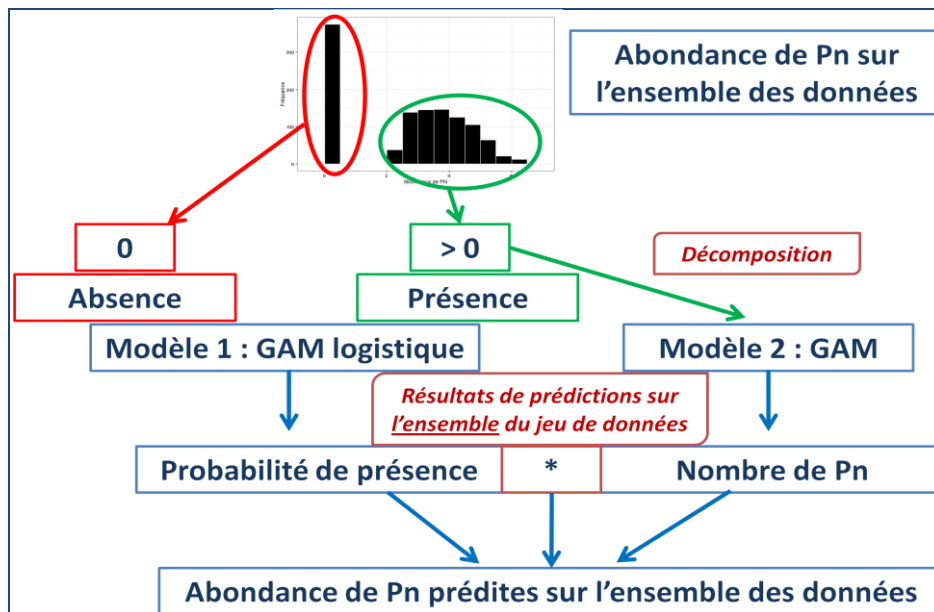


Figure 4 : Schéma de l'approche multivariée expliquant l'abondance de Pn

Les **hypothèses** posées pour ces modèles sont :

- Les observations sont **indépendantes et identiquement distribuées**.
- Il y a **homogénéité entre les lieux** des effets des différentes variables.

La **construction des modèles** est réalisée étape par étape.

Les variables explicatives sont **classées** grâce à l'analyse bivariée, par ordre décroissant de la variance expliquée. Ces variables sont introduites **une à une** dans le modèle selon cet ordre. Les variables très corrélées sont évitées en utilisant la matrice des corrélations de Spearman (Annexe 6). À chaque introduction de variable, un **test du rapport de log vraisemblance sur modèles emboîtés** est réalisé. Ce test compare deux modèles emboîtésⁱ selon leur adéquation globale aux données. S'il est significatif, alors le modèle le plus complet (avec le plus de variables) est meilleur que l'autre pour expliquer les données. Le **critère d'information d'Akaike (AIC)** est aussi observé. Il pénalise les modèles selon le nombre de variables explicatives. Si le test du rapport de log vraisemblance est significatif au risque 10% ou que l'AIC diminue, la variable est conservée dans le modèle. Les **interactions d'ordre 2** sont testées à chaque introduction de variable. « I : J » signifie l'interaction entre les variables I et J. Les interactions d'ordre supérieur ne sont pas testées pour éviter de trop complexifier les modèles.

ⁱ Un modèle comportant x variables et l'autre comportant les x variables +1 variable à tester.

Ensuite les échantillons les plus **influent**s (modifiant le plus le modèle) sont identifiés grâce à la **distance de Cook**. Les échantillons influents et atypiques sont retirés du modèle pour améliorer la modélisation sur les autres échantillons.

Le modèle final est alors **décrit** (coefficient, variables et test sur ses variables et paramètres). Lorsqu'une hypothèse est testée, la valeur de **p** correspond à la **plus petite valeur de significativité** du test. Lorsque celui-ci est inférieur à 0,05 alors le test est significatif au risque 5% et l'hypothèse testée, appelée hypothèse nulle, est rejetée. Un indicateur de significativité permet de visualiser rapidement les éléments les plus significatifs selon les tests réalisés : « *** » lorsque **p** est inférieur à 1.10^{-3} , « ** » lorsqu'il est inférieur à 1.10^{-2} , « * » en dessous de 0,05 (risque de 5%), « . » en dessous de 0,1 et un espace vide sinon.

Il est ensuite **validé** par l'analyse des **résidus** des modèles, l'analyse de la **capacité de prédiction** et la vérification de l'**effet lieu**. L'effet lieu est vérifié de deux façons complémentaires. Il est testé en effet aléatoire, vérifiant s'il existe une corrélation significative entre les échantillons d'un même lieu. Cet effet est également vérifié en effectuant une validation croisée. Le modèle est construit sur les données d'OL (échantillons d'apprentissage) puis testé sur celles de MeR. Si la prédiction du modèle est bonne et que l'effet aléatoire du lieu n'est pas significatif, le modèle est valide sur l'ensemble du Mor Braz.

III.5. Logiciels utilisés

Le logiciel utilisé pour la mise en forme des données, les réalisations graphiques et le traitement des données est le logiciel **R** (version 2.15.3) avec son interface RStudio (version 0.97.336).

IV. RÉSULTATS

IV.1. Réalisation de la base de données exploitable

IV.1.1. Mises en forme des données brutes

IV.1.1.1. Mise en forme des données du RÉPHY

De l'importation des données du RÉPHY effectuée via Quadridge² résulte une table en format csv (les champs étant séparés par un point-virgule). Chaque ligne de cette table correspond à une date, une heure, un lieu, un niveau de profondeur et un paramètre mesuré (abondances de flore totale ou de flore partielle, température de l'eau, etc...). Toutes ces informations correspondent donc à une **clé primaire** et sont nécessaires afin de déterminer une ligne précise de la table. Pour les flores, chaque ligne correspond aussi à un taxon donné. Les variables en colonne sont les suivantes :

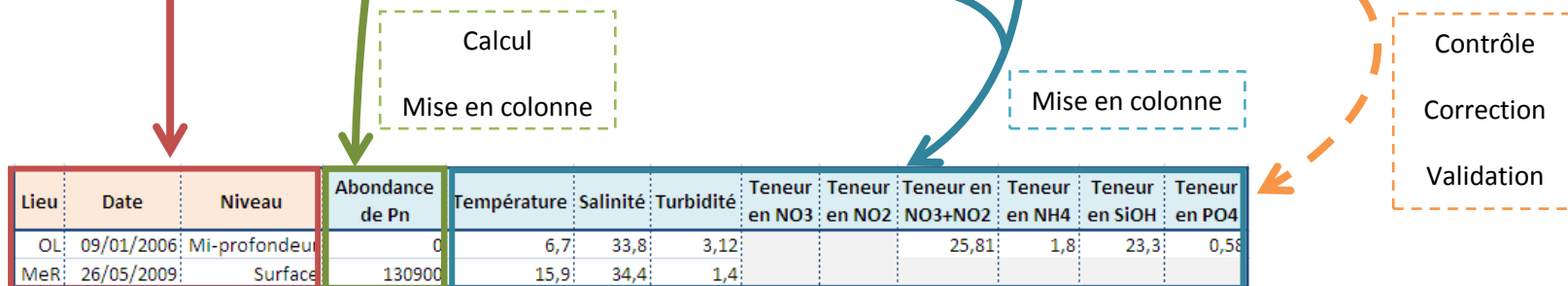
- la valeur obtenue pour le paramètre indiqué ;
- la méthode de mesure utilisée ;
- l'unité de mesure ;
- la date de validation de la donnée si elle existe ;
- et le niveau de qualité de la donnée.

Les données de cette table ont été tout d'abord **contrôlées** et **corrigées**. L'élimination des doublons ou « faux doublons » (résultant de double saisie avec différents arrondis pour les valeurs, de changement de méthode de saisie ou encore d'utilisation de plusieurs méthodes de mesure pour un même paramètre) a été effectuée en partie grâce au caractère validé ou non des données. La correction des valeurs aberrantes a nécessité la vérification de la notation sur papier et de la saisie. Les corrections relatives aux problèmes de changement de méthodes de mesure ou de défaillance de matériel ont été effectuées. Toutes ces **erreurs** aperçues sont **répertoriées** dans une table afin de faire une correction en amont dans la saisie sur Quadridge².

Puis une **mise en forme** a été réalisée de sorte que la clé primaire soit constituée de la **date**, du **lieu** et du **niveau de profondeur**. Ainsi chaque ligne correspond à un échantillon donné du RÉPHY. Les valeurs des paramètres mesurés sont mises en colonne. Au sein des paramètres flore totale, flore indicatrice et flore toxique, seule la valeur d'**abondance** des taxons correspondant au **genre Pn** est conservée. Lorsque plusieurs taxons d'un même échantillon correspondent à ce genre, ces valeurs d'abondance sont additionnées. S'il n'y en a aucun, alors cette abondance est marquée comme étant égale à zéro. Cette abondance mise en colonne est alors la variable « **Abondance de Pn** ». Un schéma de cette mise en forme des données du RÉPHY a été réalisé (Figure 5).

Tableau importé via Quadrigé²

Lieu de surveillance : Libellé	Passage : Date	Prélèvement : Niveau	Résultat : Code paramètre	Résultat : Nom du taxon	Résultat : Valeur de la mesure	Résultat : Libellé méthode	Résultat : Symbole unité de mesure	Résultat : Date de validation	Résultat : Niveau de qualité
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Melosira	200	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Protoperidinium + Peridinium	200	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Thalassiosira + Porosira	400	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Chaetoceros	500	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Nitzschia longissima	1000	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	FLORTOT	Skeletonema costatum	1100	Comptage cellules au microscope - eau	l-1	10/05/2007	Non qualifié
Ouest Loscolo	09/01/2006	Mi-profondeur	NH4		1,8	Spectrophotométrie manuelle	µmol.l-1	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	NO3+NO2		25,81	Spectrophotométrie flux	µmol.l-1	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	PO4		0,58	Spectrophotométrie manuelle	µmol.l-1	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	SALI		33,8	Capteur de conductivité in situ	sans unité	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	SiOH		23,3	Spectrophotométrie manuelle	µmol.l-1	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	TEMP		6,7	Capteur de température in situ	°C	10/05/2007	Bon
Ouest Loscolo	09/01/2006	Mi-profondeur	TURB		3,12	Turbidimètre optique dans échantillon	NTU	10/05/2007	Bon
Men er Roue	26/05/2009	Surface (0-1m)	FLORIND	Alexandrium	700	Comptage cellules au microscope - eau	l-1	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	FLORIND	Pseudo-nitzschia	4600	Comptage cellules au microscope - eau	l-1	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	FLORIND	Pseudo-nitzschia, groupe des effilées	6300	Comptage cellules au microscope - eau	l-1	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	FLORIND	Pseudo-nitzschia, groupe des fines	120000	Comptage cellules au microscope - eau	l-1	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	SALI		34,4	Capteur de conductivité in situ	sans unité	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	TEMP		15,9	Capteur de température in situ	°C	05/03/2012	Non qualifié
Men er Roue	26/05/2009	Surface (0-1m)	TURB-FNU		1,4	Turbidimètre optique dans échantillon	FNU	05/03/2012	Non qualifié



Les colonnes en orange déterminent une ligne du tableau (clé primaire).

Tableau mis en forme

Figure 5 : Exemple de mise en forme des données du RÉPHY

Enfin les **données non exploitables** ont été **éliminées**. L'identification du phytoplancton Pn a évolué au cours du temps et, afin d'éviter des erreurs d'abondance de Pn dues à cette évolution, la période analysée se restreint à la période comprise entre **1995**, année à partir de laquelle cette identification s'est stabilisée, et 2012.

Les niveaux de profondeur conservés dans la table sont la **mi-profondeur** et la **surface**. Les autres niveaux de prélèvement ont un faible effectif et sont ponctuels dans le temps. Au final, pour chaque date, seul un niveau de profondeur est renseigné.

Les méthodes de mesure de teneur en chlorophylle a dans le cadre du RÉPHY ont changé au cours du temps, provoquant des incertitudes dans les données. Ces données n'ont pas été conservées au profit des données par satellite.

Afin de réduire les valeurs manquantes pour les nutriments azotés, les variables NO₂, NO₃ et NH₄ ou NO₂+NO₃ et NH₄ sont sommés. Ainsi, la variable résultante est la teneur en **azote inorganique** notée **N**. Parmi les nutriments azotés, seule la variable N est conservée dans la base.

La base résultant de cette mise en forme contient **1172 lignes**, avec une **bonne répartition entre OL et MeR**. 583 sont des échantillons d'OL et 589 sont des échantillons de MeR, prélevés entre **1995 et 2012**.

IV.1.1.2. Mise en forme des données hors RÉPHY

Les **données** hors RÉPHY sont également mises en forme.

Les données concernant la température par satellite, la teneur en matière inorganique en suspension, la teneur en chlorophylle a, le coefficient d'extinction lumineuse et le gradient d'énergie de stratification concernent toute la zone du Mor Braz découpée en maille jusqu'à 1km² de surface. Afin de permettre leur exploitation et leur ajout à la base précédente, les valeurs de chacun de ces paramètres sont **moyennées** pour l'ensemble des mailles situées dans la **baie de Quiberon** et séparément dans la **baie de Vilaine**ⁱ. Ainsi, on dispose de données journalières sur ces deux baies pour ces différents paramètres.

Ces données ont été **contrôlées**. Les données de **teneur en chlorophylle a**, de **coefficient d'extinction lumineuse** et de **gradient d'énergie de stratification journalières** ont été **jointes** à la base précédente selon les **dates** et les lieux (**OL** est associé à la baie de Vilaine et **MeR** à la baie de Quiberon). Les données journalières de température par modèle (Eco Mars 3D) et de teneur en mics ne sont pas jointes à la table (redondance d'informations avec respectivement la température mesurée par le RÉPHY et la turbidité).

ⁱ Ce travail a été réalisé par Jean-Yves Stanisière, ingénieur de l'Ifremer à la station de la Trinité-sur-Mer.

Les données de rayonnement récupérées sur plusieurs stations météorologiques sont **moyennées** par date de façon à obtenir une variable de **rayonnement journalier dans le Mor Braz**. Cette variable ainsi que les **données journalières de débits de la Loire et de la Vilaine** sont ensuite jointes à la base selon la **date**.

IV.1.2. Les variables calculées et ajoutées à la base

Quatre types de données supplémentaires ont été calculés puis ajoutés à la base.

IV.1.2.1. Les indices écologiques

Pn est un phytoplancton vivant parmi d'autres et les **interactions** avec **les autres espèces phytoplanctoniques** existent (Trainer et al., 2012 ; Downes-Tettmar et al., 2013). Afin de prendre en compte ces interactions d'une façon simplifiée, des indices écologiques ont été étudiés puis calculés. Les indices choisis sont l'**équitabilité de Piélou** (notée « **J** ») et la **richesse** (notée « **Hmax** »). Ils dépendent d'un niveau taxonomique choisi, qui est souvent l'espèce. La richesse représente la **biodiversité** de la **population phytoplanctonique** (Hmax augmente avec le nombre d'espèces différentes). L'équitabilité de Piélou représente un certain **équilibre** de cette même population au niveau des **proportions** de chaque espèce. J est maximal quand il est égal à Hmax, c'est-à-dire quand toutes les proportions pour chaque espèce sont égales (population équilibrée). Il diminue avec les déséquilibres dans ces proportions donc avec l'apparition de dominances/d'espèces en plus grande proportion que d'autres.

Le calcul de ces indices a été réalisé uniquement à partir de l'ensemble des données brutes de **flore totale**. Il a requis la **classification** phylogénétique (classification permettant de marquer les degrés de parenté entre espèces) de **tous les taxons** rencontrés dans les données de l'étude. Celle-ci a été effectuée à partir du **WORMS** : WOrld Register of Marine Species, registre répertoriant toutes les espèces marines rencontrées dans le monde et la classification associée. Un exemple de classification est disponible en annexe.

Afin de choisir un **niveau taxonomique de référence** pour le calcul de ces indices écologiques, la **qualité d'identification** des cellules de phytoplancton des échantillons a été analysée. Cette analyse a montré que les techniques d'identification ne permettent de prendre l'espèce comme référence. Le niveau taxonomique choisi suite à cette analyse est l'**ordre**. Le calcul de ces indices écologiques est détaillé en Annexe 4.2.

IV.1.2.2. Les rapports de nutriments

Les **rapports de nutriments** ont une influence sur Pn (Trainer et al., 2012). Les rapports des teneurs en azote/phosphates noté **N/P**, silicates/azote noté **Si/N** et phosphates/silicates noté **P/Si** ont donc été calculés à partir des données brutes de nutriments du RÉPHY et ajoutés à la base.

IV.1.2.3. L'énergie utile

Le rapport **rayonnement sur coefficient d'extinction lumineuse** permet d'obtenir un indicateur de l'« **énergie utile** » : c'est l'énergie lumineuse qui a pu pénétrer dans l'eau et peut être utilisé par le phytoplancton. Ce rapport est également calculé puis ajouté à la base.

IV.1.2.4. L'effet moyenné sur la semaine précédente

L'effet des variables environnementales sur l'abondance de Pn n'est **pas immédiat**. L'abondance de Pn mesurée à une date précise peut résulter des conditions précédant cette date sur une certaine période. Ceci a pu être pris en compte pour les variables **non originaires du RÉPHY**, dont les données sont **journalières**. Pour chacune de ces variables et pour chaque date donnée, un effet moyen sur une semaine est calculé en moyennant les valeurs de la date donnée et des **6 jours précédents** cette date. Les variables **température moyenne, teneur en mies moyenne, teneur en chlorophylle a moyenne, coefficient d'extinction lumineuse moyen, stratification moyenne, rayonnement moyen, énergie utile moyenne** et **débites de la Loire et de la Vilaine moyens** ont donc été intégrées à la base. Excepté pour la température par modèle et la teneur en mies, les données journalières sont conservées.

Un schéma récapitule l'ensemble des actions mises en œuvre pour mettre en forme et obtenir la base exploitable à partir des données brutes (Figure 6). Au final, la base dispose de **28 variables environnementales** quantitatives explicatives (Figure 6). Cette base est ensuite exploitée afin de répondre à l'objectif de l'étude, qui est d'expliquer l'abondance de Pn et ses efflorescences en fonction des conditions environnementales dans le Mor Braz.

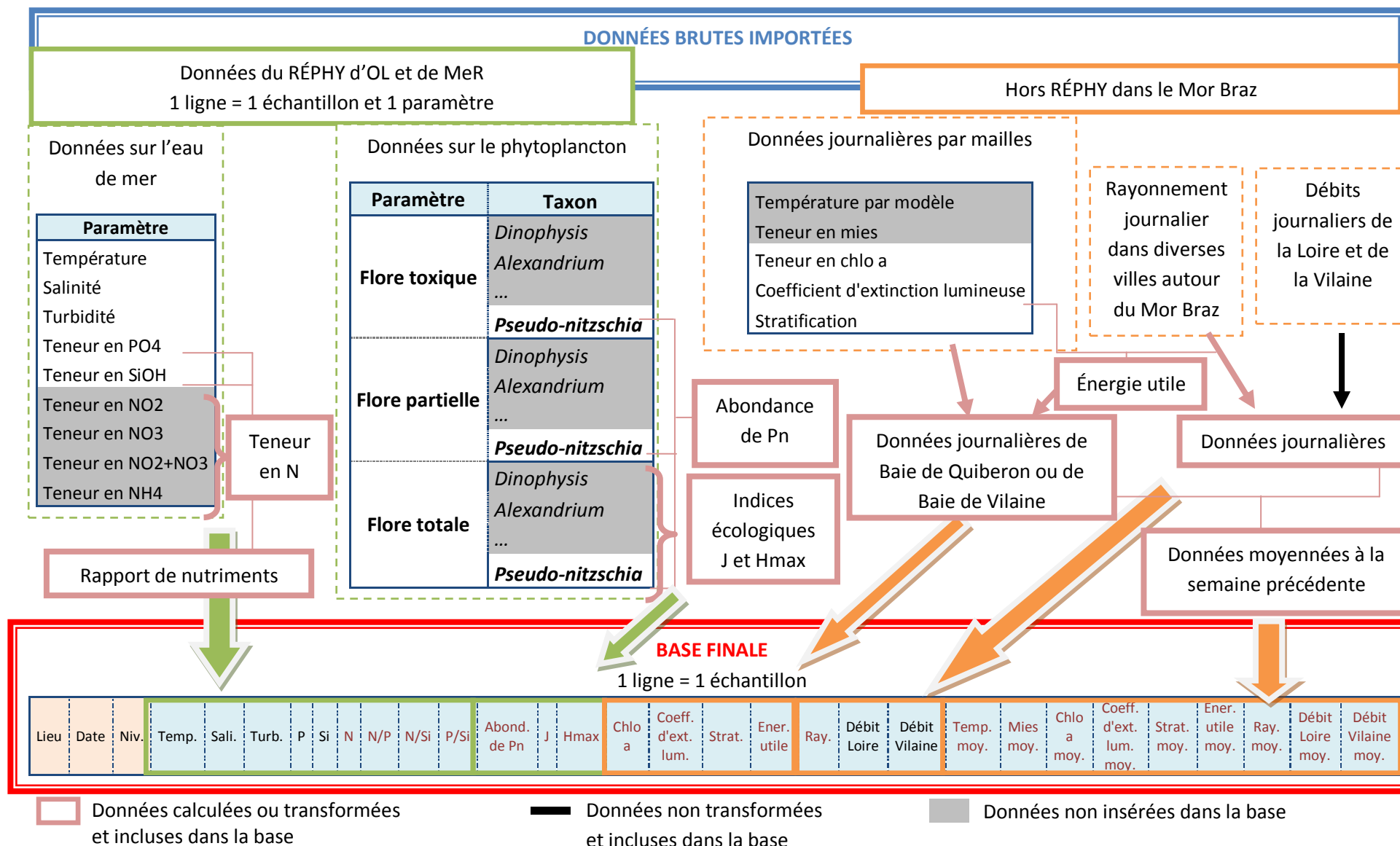


Figure 6 : Construction de la base de données exploitée

IV.2. Description de l'abondance de Pn

IV.2.1. Visualisation graphique

L'abondance de Pn varie de 0 à 2,5 millions de cellules/L. On rappelle que celle-ci est transformée en log₁₀. La

Figure 7 représente l'abondance de Pn en fonction de la date à OL de 2000 à 2005 (la période est restreinte pour faciliter la lecture du graphe). Elle illustre l'apparition régulière de fortes abondances dépassant les 100 000 cellules/L (en rouge). De telles efflorescences provoquent une alerte avec le comptage de flore toxique et la surveillance du taux d'acide domoïque dans les coquillages. Lorsque l'abondance est faible (en dessous de 10 000 cellules/L), celle-ci est représentée en vert (orange sinon).

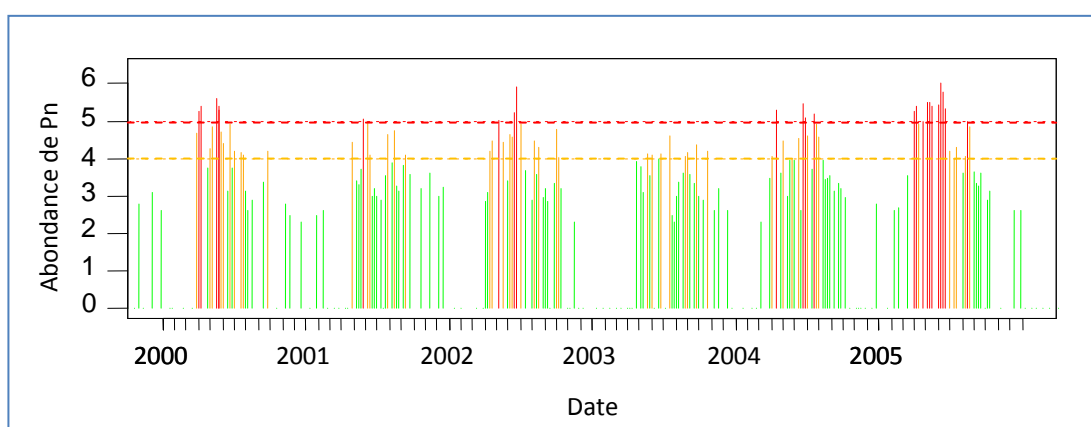


Figure 7 : L'abondance de Pn à OL de 2000 à 2005

IV.2.2. Saisonnalité et tendance

Le GAM permet d'observer la variation temporelle de l'abondance de Pn sur une année et le long des années en prenant en compte toute la période d'étude. Les résultats sont présentés en Figure 8 et Figure 9.

Les courbes de lissage réalisées par le GAM sont représentées en noir. Les courbes en pointillées correspondent à l'intervalle de confiance (à 95%) de la courbe. Le nuage de points des valeurs d'abondance de Pn en fonction du jour de l'année ou de l'année est représenté en bleu.

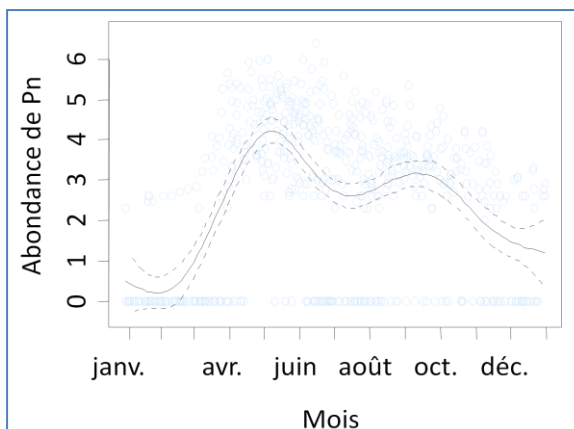


Figure 8 : Saisonnalité de l'abondance de Pn à Ouest Loscolo

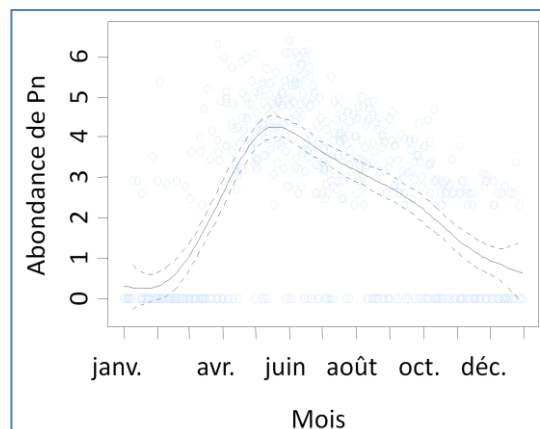


Figure 9 : Saisonnalité de l'abondance de Pn à Men er Roue

La saisonnalité de l'abondance de Pn est clairement mise en évidence. Les **saisonnalités** sont **différentes entre MeR et OL**. Deux pics sont observés à OL contre un seul à MeR. La période de croissance de Pn est comprise entre **mars et novembre** (du printemps à l'automne). Son abondance la plus forte dans l'année est atteinte globalement en mai. A MeR celle-ci décroît ensuite globalement, alors qu'à OL un pic apparaît vers les mois de septembre/octobre. On remarque que cette moyenne saisonnière est plutôt faible (jusqu'à 40 000 cellules/L).

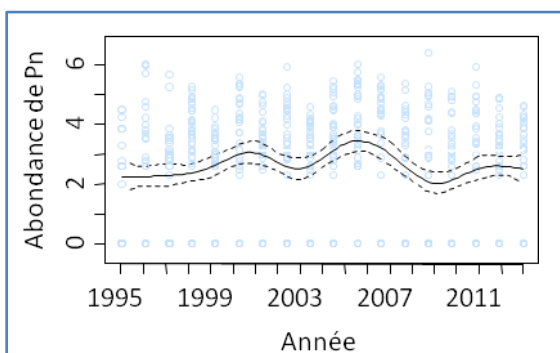


Figure 10 : Tendence de l'abondance de Pn à OL

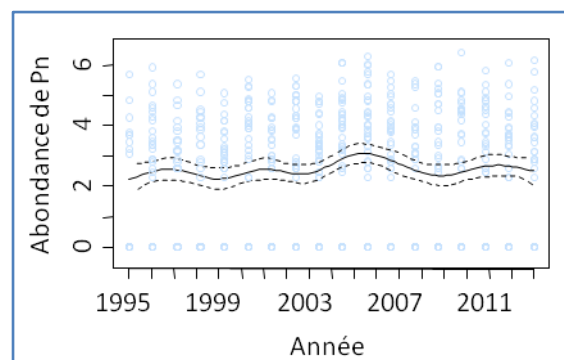


Figure 11 : Tendence de l'abondance de Pn à MeR

Des **variations inter-annuelles** sont mises en évidence (*Figure 10* et *Figure 11*) mais globalement l'abondance de Pn est stable sur la période entre 1995 et 2012 (pas de tendance à l'augmentation). Les **tendances** de l'abondance de Pn ne diffèrent pas entre OL et MeR.

IV.2.3. Analyse des efflorescences de Pn à OL et MeR

L'analyse des efflorescences a permis de **décrire les efflorescences de Pn** observés de 1995 à 2012 à OL et MeR.

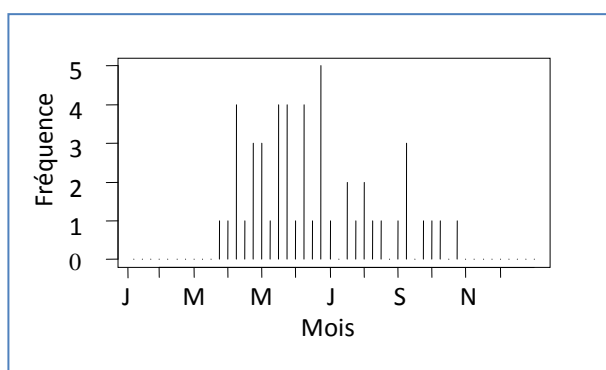


Figure 12 : Fréquence des efflorescences en fonction du quart de mois de l'année à OL

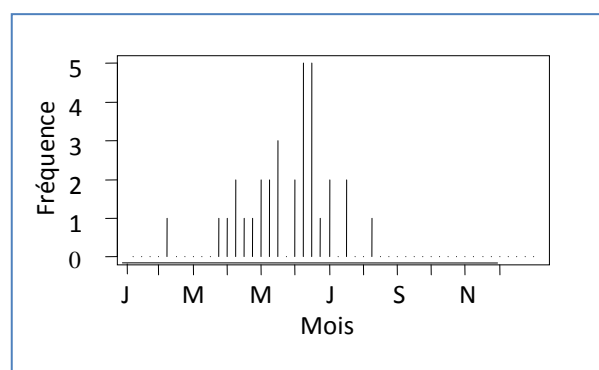


Figure 13 : Fréquence des efflorescences en fonction du quart de mois de l'année à MeR

Les efflorescences apparaissent le plus fréquemment aux mois d'avril et de mai (Figure 12 et Figure 13). On remarque une différence importante entre OL et MeR. Des pics d'abondance apparaissent encore en automne à OL (fin août et septembre) alors qu'il n'y en a plus après juillet à MeR. Cela concorde avec ce qui a été vu dans les Figures 7 et 8.

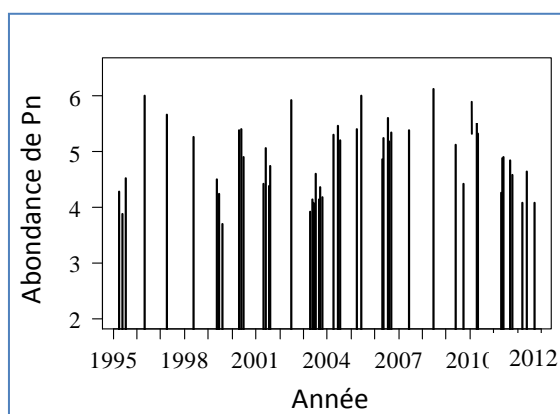


Figure 14 : Intensité et dates des efflorescences de Pn le long de la période d'étude à OL

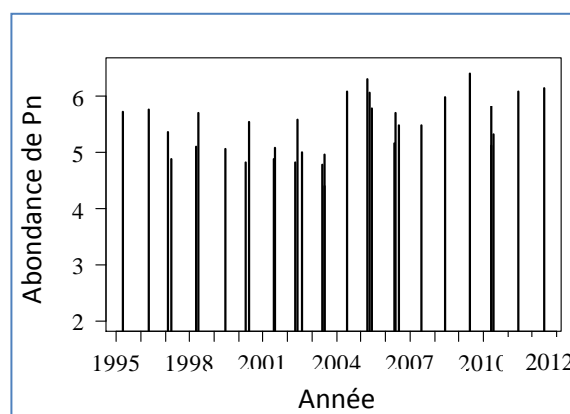


Figure 15 : Intensité et dates des efflorescences de Pn le long de la période d'étude à MeR

Le **nombre d'efflorescences** est **plus élevé à OL** qu'à MeR (Figure 14 et Figure 15). 49 pics sont observés à OL contre 32 à MeR. L'**intensité** des efflorescences **varie** d'une année à l'autre. Par exemple à OL, l'année 2003 comporte 8 petites efflorescences de Pn de 10 000 à 30 000 cellules/L d'eau de mer environ, alors que l'année 2006 comporte 5 efflorescences de 100 000 à 300 000 cellules/L environ (ce qui correspond à un dépassement du seuil d'alerte fixé par le RÉPHY).

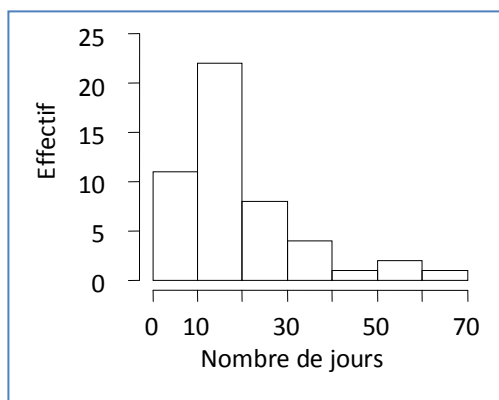


Figure 16 : Fréquence de la durée des efflorescences de Pn en nombre de jours à OL

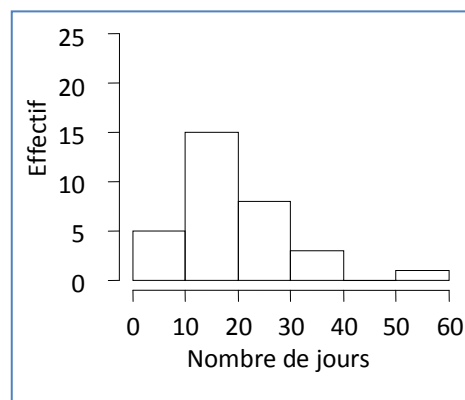


Figure 17 : Fréquence de la durée des efflorescences de Pn en nombre de jours à MeR

La durée d'efflorescences la plus fréquemment observée est comprise entre 10 et 20 jours, que ce soit à OL ou à MeR (Figure 16 et Figure 17). Les durées maximales se situent aux alentours de deux mois : trois cas observés à OL et un à MeR.

Les pics d'abondance de Pn sont des phénomènes **réguliers** mais d'**intensité** et de **durée variables**. Des **différences** existent entre **OL** et **MeR** au niveau de la **variation temporelle de l'abondance de Pn** et de **l'apparition de ces efflorescences**. Nous allons maintenant tenter d'expliquer les variations d'abondance de Pn par les variations des conditions environnementales grâce à la modélisation. Pour cela, des analyses sont effectuées au préalable afin de modéliser au mieux cette relation.

IV.3. Analyse bivariée

IV.3.1. Vérification de la linéarité des logits

Une **représentation graphique des logits** pour chaque variable a permis de constater leur **non linéarité**. Une partie de ces représentations graphiques avec explication de la méthode de réalisation de ces graphes est disponible en Annexe 7. Cela justifie le choix du GAM plutôt que le GLM pour expliquer l'absence ou la présence de Pn.

En ce qui concerne la température et la chlorophylle a, les logits peuvent être approchés à une relation linéaire. Une augmentation d'une unité de **température** ou de **chlorophylle a** dans l'eau augmente d'environ **25 à 26 %** le risque d'apparition de Pn, en ne prenant qu'un seul effet à la fois (Rapport des risques, appelé Odds Ratio non ajusté et calculé à partir des coefficients des régressions logistiques univariées expliquant l'absence/présence de Pn en fonction de la température ou de la chlorophylle a).

IV.3.2. Type de relation entre l'abondance de Pn et les variables environnementales

L'analyse de ces relations est une étude préalable au modèle multivarié expliquant l'abondance de Pn en cas de présence de Pn. Ces relations sont étudiées sur les données où Pn est présente. L'analyse bivariée met en évidence que les **relations entre l'abondance de Pn** et chacune des **variables environnementales** ne sont **pas linéaires**. Deux exemples sont présentés.

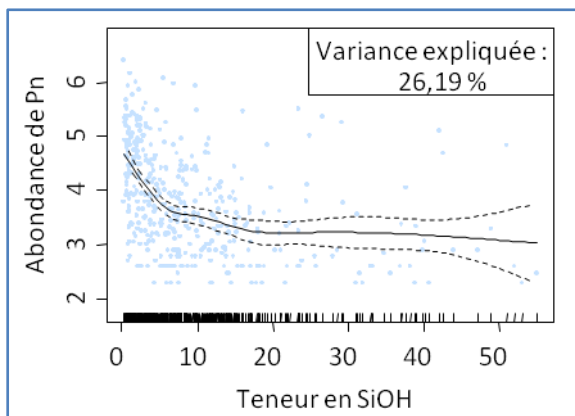


Figure 18 : L'abondance de Pn en fonction de la teneur en SiOH dans l'eau de mer

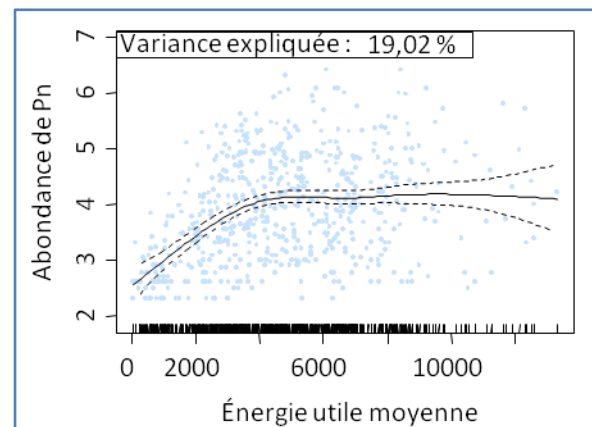


Figure 19 : L'abondance de Pn en fonction de l'énergie utile moyenne

Les courbes de lissage de l'abondance de Pn en fonction de la teneur en SiOH (Figure 18) et de l'énergie utile moyenne (Figure 19) sont monotones (respectivement croissante ou décroissante) puis constantes à partir d'un seuil. L'abondance de Pn décroît avec la teneur en SiOH jusqu'à un seuil aux alentours de 18 $\mu\text{g/L}$. Passé ce seuil, l'abondance de Pn est constante entre 1000 et 10 000 cellules/L. Au contraire l'abondance de Pn croît avec l'énergie utile moyenne puis passé un seuil de 2500 J/cm^2 environ, l'abondance de Pn reste stable à une valeur globale de 10 000 cellules/L.

La teneur en SiOH explique la variance de l'abondance de Pn à 26,19 % en analyse bivariée contre 19,02 % pour l'énergie utile moyenne (Figure 18 et Figure 19). Ces deux variables font parti des premières variables testées au cours de la construction du modèle multivarié expliquant la croissance de Pn.

IV.4. Les modèles explicatifs et prédictifs

Le jeu des données traitées a été **décomposé**. **47%** des données d'abondance de Pn sont des **zéros**, soit d'absence de Pn (Figure 20). Cela conforte l'approche effectuée, qui est donc tout d'abord d'expliquer par un premier modèle l'**apparition** de Pn (l'absence ou la présence), puis par un deuxième modèle d'expliquer la **croissance** de Pn (l'abondance de Pn lorsqu'elle est présente).

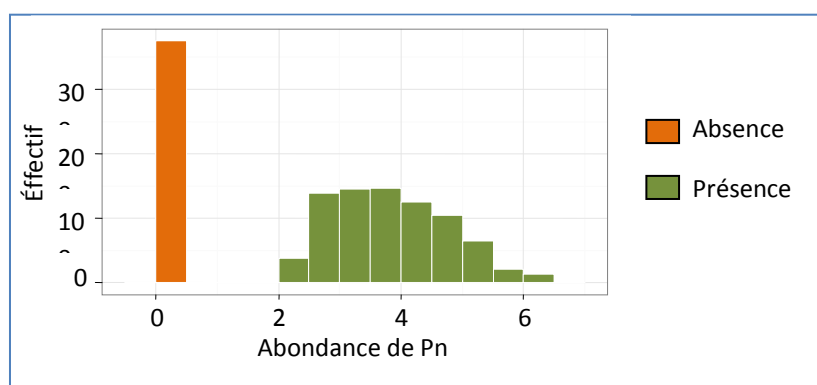


Figure 20 : Histogramme de l'abondance de PN sur l'ensemble des données

IV.4.1. Modèle expliquant l'apparition

Six variables explicatives ont été sélectionnées dans le **GAM logistique binaire** expliquant l'absence ou la présence de Pn : la teneur en SiOH, la température, Hmax (la richesse), J (l'équitabilité de Piélou), la teneur en chlo a et le débit de la Vilaine. Suite à cela, l'analyse de la distance de Cook a permis d'améliorer la capacité de prédiction globale du modèle en retirant **deux échantillons** du jeu de données pour la construction du modèle (échantillons influents et atypiques où Pn est présent en hiverⁱ). Ce modèle a été construit sur 480 données et explique 54% de la variance du logit.

Le Tableau 1 présente le **GAM logistique** final avec les informations relatives à la **significativité** de l'ordonnée à l'origine (seul paramètre quantifiable) et des variables (ou plus précisément du lissage réalisé entre la variable explicative et la variable à expliquer). Chacun de ces

ⁱ Les échantillons du 19/12/2009 à OL et du 27/12/2006 à MeR où l'abondance de Pn est de 200 cellules/L

tests sont basés sur le principe du test de Wald et teste la nullité des coefficients (pour les lissages, c'est l'espérance de l'ensemble des coefficients qui est testée).

Tableau 1 : Variables explicatives et coefficients associés du modèle de GAM logistique obtenu

	Coefficient estimé	Ecart-type	Valeur de p	Significativité
Ordonnée à l'origine	1,27	1,95E-01	9,34E-11	***
Teneur en SiOH			1,60E-05	***
Hmax			3,44E-05	***
J			5,57E-03	**
Teneur en chlo a			1,50E-03	**
Température			1,33E-02	*
Débit de la Vilaine			1,14E-01	.
Température : J			9,32E-02	
SiOH : Débit de la Vilaine			1,33E-01	

Tous ces tests sont **significatifs** au risque 5 % ($p < 0,05$), exceptés pour le débit de la Vilaine et les interactions entre la température et J et entre la teneur en SiOH et le débit de la Vilaine. Ces trois derniers éléments du modèle sont cependant conservés car les tests du rapport de log vraisemblance sur modèles emboîtés sont significatifs au risque 5% (de plus l'AIC diminue).

L'équation de ce modèle est donc :

$$\ln\left(\frac{P}{1-P}\right) = 1,27 \pm 0,38 + f_{\text{Teneur en SiOH}}(\text{Teneur en SiOH}) + f_{\text{Hmax}}(\text{Hmax}) + f_J(\text{J}) + f_{\text{Teneur en chlo a}}(\text{Teneur en chlo a}) + f_{\text{Température}}(\text{Température}) + f_{\text{Débit de la Vilaine}}(\text{Débit de la Vilaine}) + f_{\text{Température : J}}(\text{Température : J}) + f_{\text{Teneur en SiOH : Débit de la Vilaine}}(\text{Teneur en SiOH : Débit de la Vilaine})$$

Ce modèle explicatif est aussi **prédictif**. Il donne des probabilités de présence de Pn en fonction des variables environnementales. La *Figure 21* représente ces probabilités données par le modèle en fonction de l'absence (notée 0) ou de la présence (notée 1) réelle de Pn. On remarque qu'un **seuil de 0,5** permet d'assez bien différencier l'absence de la présence grâce au modèle. Lorsque le modèle donne une probabilité de présence inférieure à 0,5, on prédit alors l'absence et inversement la présence.

Le Tableau 2 confronte ce qui est prédit par le modèle et la réalité en prenant ce seuil. **76,6%** des absences de Pn sont bien prédites par le modèle : c'est la **spécificité**, soit la probabilité de prédire l'absence en cas d'absence. **90,4%** des présences de Pn sont bien prédites par le modèle : c'est la **sensibilité**, soit la probabilité de prédire la présence en cas de présence. L'**erreur de classement** est de **14,5%**.

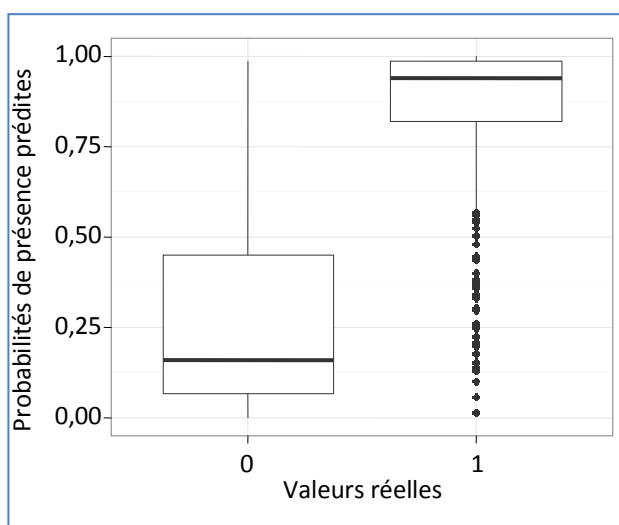


Tableau 2 : Matrice de confusion du modèle en prenant un seuil de 0,5

		Réel	
		Absence	Présence
Prédit	Absence	131	30
	Présence	40	281

Figure 21 : Probabilité de présence de Pn données par le GAM logistique en fonction de la présence ou absence de Pn réelle

La **courbe ROC** représente la spécificité et la sensibilité pour des valeurs régulières de seuil entre 0 et 1 (Figure 22). Les intervalles de confiance (en bleu) sont créés par bootstrap (méthode qui consiste à tester les prédictions sur des échantillons tirés sans remise n fois). Par exemple, un seuil de 0 signifie que lorsque la probabilité du modèle est inférieure à 0, on prédit que Pn est absente donc Pn est toujours prédite présente. La sensibilité est donc dans ce cas de 100 % tandis que la spécificité est de 0 %. Elle permet d'apprécier la capacité prédictive du modèle.

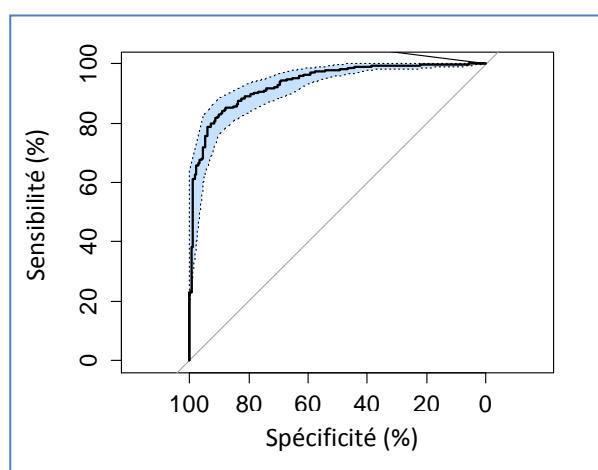


Figure 22 : Courbe ROC du GAM logistique obtenu

L'**aire** sous cette courbe est de **0,93** : cela signifie que le modèle a une bonne capacité prédictive (pour une capacité prédictive maximale, l'aire est de 1).

Afin de **valider** le modèle, les résidus de Pearson sont analysés et l'effet aléatoire du lieu est testé.

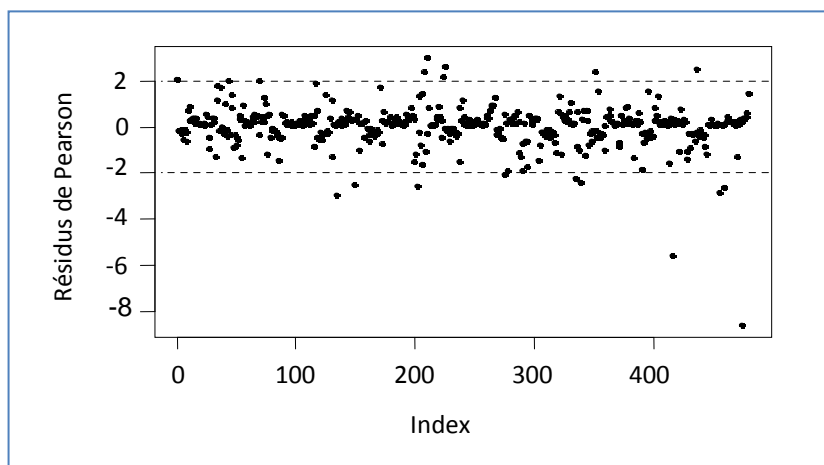


Figure 23 : Résidus de Pearson des données modélisés par le GAM logistique

Seuls 2,5 % des points de ce graphe (Figure 23) sont en dehors de l'intervalle $[-2,2]$. La modélisation du GAM logistique est donc bonne. Aucune corrélation entre les résidus n'est apparente. Les résidus de Pearson révèlent cependant deux points particulièrement mal modélisés. Ils correspondent aux échantillons du 02/08/2010 à OL et du 21/11/2011 à MeR, où Pn est absente alors que le modèle donne une probabilité de présence très élevée (entre 0,96 et 0,99).

L'effet du lieu est vérifié. Le test de l'effet aléatoire montre que la corrélation entre les échantillons d'un même lieu n'est pas significative au risque 5% ($p=0,48$). La validation croisée effectuée montre que le modèle construit sur les données d'OL prédit mieux que le hasard sur MeR (Figure 24). L'aire sous la courbe ROC est en effet de 76,1 %. Le modèle est acceptable. L'absence de Pn est mieux prédite que sa présence (Figure 25). Le modèle reste valide d'une zone à l'autre du Mor Braz.

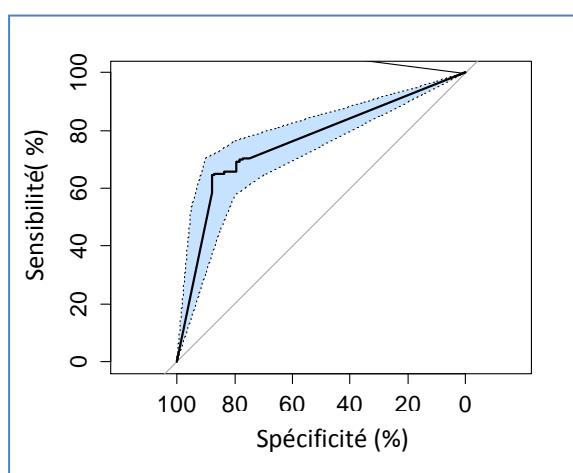


Figure 24 : Courbe ROC du GAM logistique lors de la validation croisée

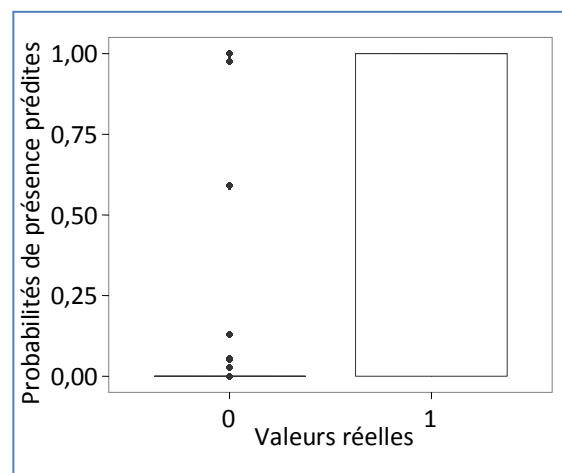


Figure 25 : Prédiction du GAM lors de la validation croisée

Ce GAM logistique est donc **correcte** (bonne adéquation aux données et bonne prédiction). Il permet de bien expliquer l'absence ou présence de Pn dans le Mor Braz en fonction des conditions environnementales. La validation croisée montre cependant que l'efficacité de ce modèle n'est pas homogène dans le Mor Braz.

IV.4.2. *Modèle expliquant la croissance de Pn*

Afin d'expliquer la croissance de Pn, un deuxième modèle de type GAM est construit. Le GAM retenu suite aux tests de distribution de l'abondance de Pn utilise la **loi normale** et la fonction de lien **identité**. **Cinq variables** sont sélectionnées dans ce modèle : la **stratification moyenne**, le **rapport P/Si**, la **teneur en SiOH**, l'**énergie utile moyenne** et J. Seule l'interaction entre J et le rapport P/Si est conservée (toujours selon le test du rapport de Vraisemblance et l'AIC). La distance de Cook révèle des points influents mais aucun ne semble atypique. Aucun échantillon n'est retiré de la construction du modèle.

Le modèle est construit sur 219 données. Le Tableau 3 présente le GAM obtenu et les tests associés à chacun de ces éléments.

Tableau 3 : Le GAM expliquant la croissance de Pn

	Coefficient estimé	Ecart-type	Valeur de p	Significativité
Ordonnée à l'origine	3,88	4,50E-02	<2E-16	***
Stratification moyenne			3,84E-04	***
P/Si			4,80E-03	**
SiOH			1,20E-02	*
Energie utile moyenne			1,46E-02	*
J			6,28E-01	
P/Si : J			2,44E-04	***

D'après ces tests, les variables du modèle sont toutes significatives ($p < 0,05$) excepté pour J. Cette dernière doit être conservée avec son interaction avec P/Si (qui est significative). L'équation du modèle obtenu est :

$$\text{Abondance de Pn} = 3,88 \pm 0,09 + f_{\text{SiOH}}(\text{SiOH}) + f_{\text{Énergie utile moyenne}}(\text{Énergie utile moyenne}) + f_{\text{P/Si}}(\text{P/Si}) + f_{\text{J}}(\text{J}) + f_{\text{Stratification moyenne}}(\text{Stratification moyenne}) + f_{\text{P/Si : J}}(\text{P/Si : J})$$

La variance expliquée par ce modèle est d'environ 57,1%.

4,11% des résidus de Pearson sont en dehors de l'intervalle [-2,2] (Figure 26). La modélisation des échantillons par le modèle obtenu est donc bonne. Des visualisations graphiques complémentaires montrent également que les résidus du modèle sont normaux et de variance

homogène (Annexe 8.2). Aucune corrélation n'est mise en évidence. La modélisation du GAM réalisé est donc bonne.

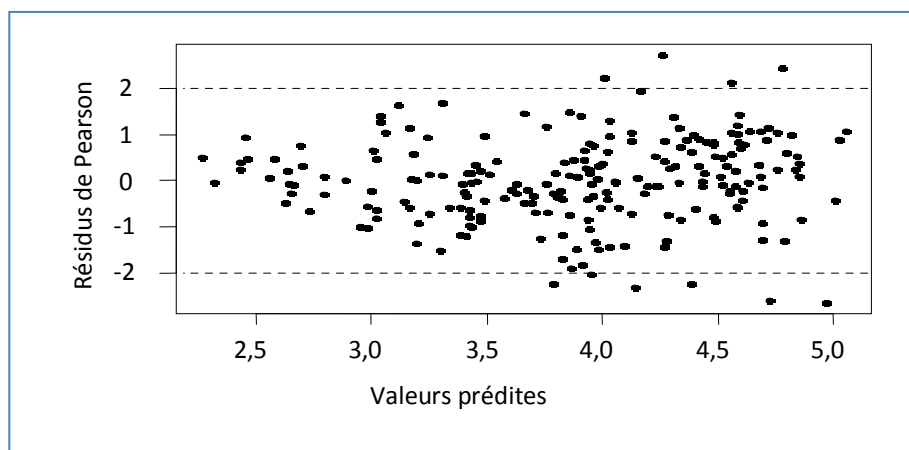


Figure 26 : Résidus de Pearson du GAM obtenu

Ce modèle prédit l'abondance de Pn en fonction de ces 5 variables explicatives. Afin de visualiser la bonne prédiction du modèle, un graphe représentant les valeurs d'abondance de Pn prédites en fonction des valeurs d'abondance de Pn réelles est réalisé (Figure 27).

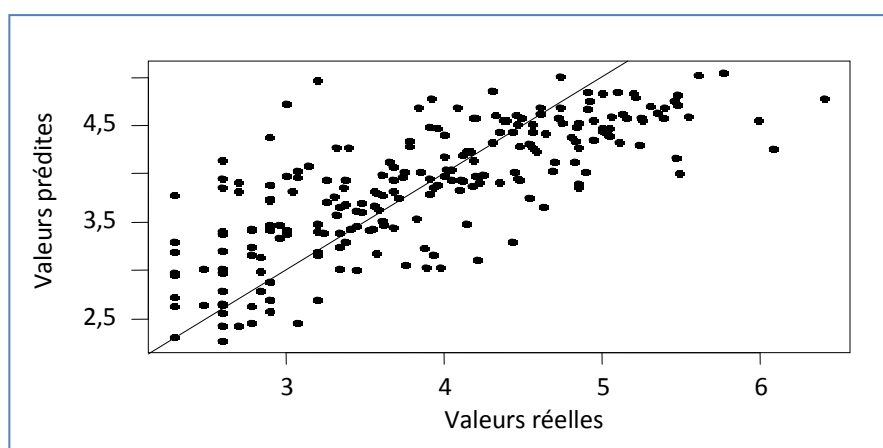


Figure 27 : Abondance de Pn prédite par le modèle en fonction de l'abondance de Pn réelle

Lorsque la prédiction est optimale, les points représentés sur ce type de graphique suivent la droite tracée en noir d'équation $y=x$. Au vu de ce graphe, la **capacité de prédiction** du modèle est très **moyenne**.

De même que pour le GAM logistique, l'**effet aléatoire** du lieu n'est **pas significatif** au risque 5% ($p=0,93$). La **validation croisée** montre cependant que la modélisation d'un lieu à un autre n'est pas parfaite (Figure 28). On note un décalage des valeurs prédites sur la gauche. Le modèle n'est donc pas capable de bien prédire l'abondance de Pn sur l'ensemble du Mor Braz.

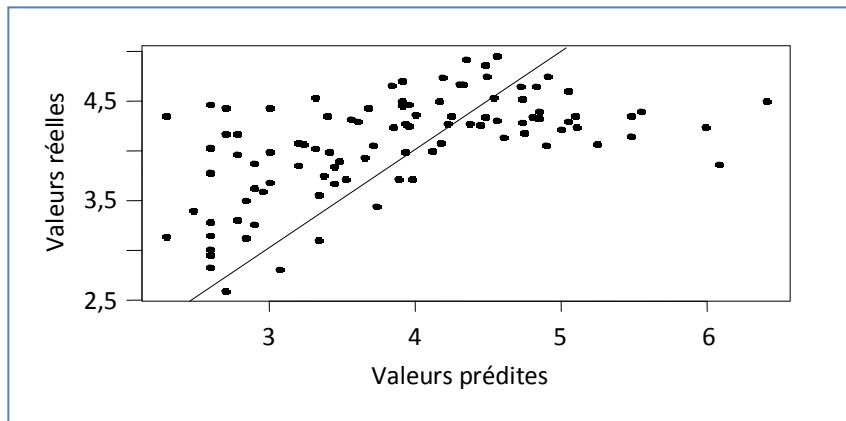


Figure 28 : Prédiction du GAM sur les données en présence de Pn suite à la validation croisée

Ce GAM expliquant la croissance de Pn dans l'ensemble du Mor Braz est acceptable (modèle valide mais capacité de prédiction moyenne).

IV.4.3. Combinaison des 2 modèles

Les 2 modèles sont ensuite **combinés**. Les prédictions d'abondance de Pn du GAM (deuxième modèle) sur l'**ensemble** des données sont « corrigées » par les probabilités de présence données par le GAM logistique (premier modèle). Ainsi le phénomène complet (apparition de Pn et des efflorescences) est expliqué.

La Figure 29 représente la prédiction du deuxième modèle sur l'ensemble des données. Elle montre que le modèle prédit assez bien l'abondance de Pn lorsqu'elle est présente mais qu'il ne prédit aucune abondance nulle. La Figure 30 représente ces mêmes prédictions multipliées par zéro quand le deuxième modèle donne une probabilité de présence inférieure à 0,5.

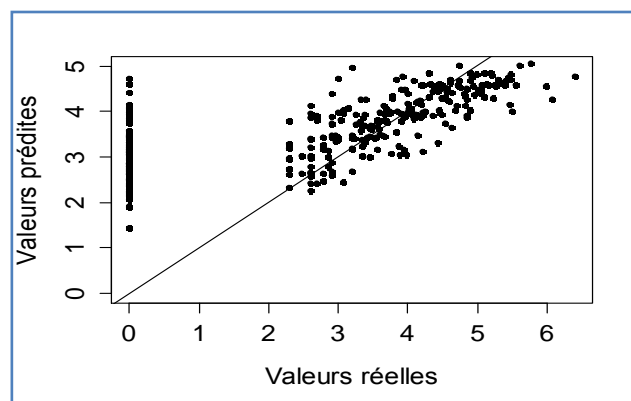


Figure 29 : Prédiction du deuxième modèle sur l'ensemble des données

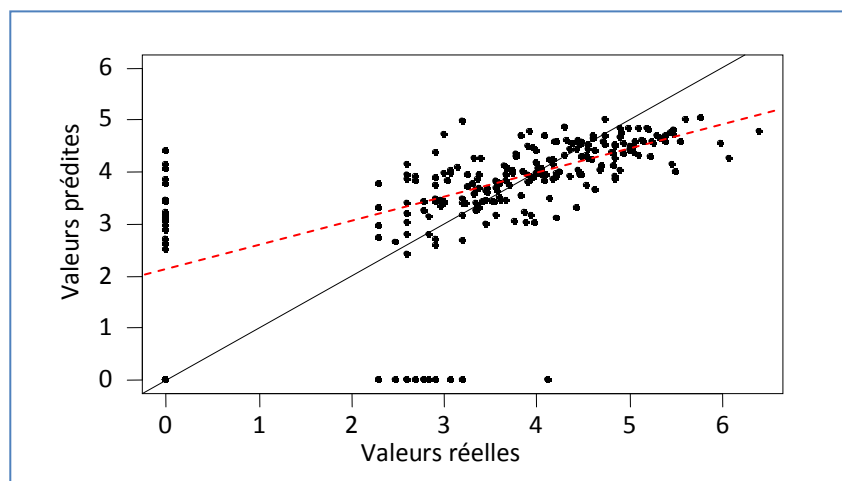


Figure 30 : Prédiction du deuxième modèle une fois combiné au premier modèle

Finalement les deux modèles combinés sont capables de prédire l'abondance de Pn sur l'ensemble des données (Figure 30). Les informations des Tableau 4 et Tableau 5 permettent d'apprécier les résultats issus de cette combinaison.

76,1% des absences de Pn sont bien prédites et 89,9% des présences sont bien prédites (Tableau 4). Parmi les faux négatifs (7,10% des échantillons), l'abondance de Pn prédite par le modèle a une valeur comprise entre 200 et 13 000 cellules/L, ce qui est bien en dessous du seuil d'alerte (100 000 cellules/L). Parmi les faux positifs (7,10% des échantillons), l'abondance de Pn prédite est comprise entre 316 et 25 500 cellules/L, ce qui est aussi bien en dessous du seuil. Il est important de noter que les efflorescences supérieures à 100 000 cellules/L (seuil d'alerte) sont mal prédites par le modèle.

Tableau 4 : Prédiction de l'absence et de la présence par le modèle combiné

	Faux positifs	Faux négatifs	Vrais négatifs	Vrais positifs	Total
Effectifs	22	22	70	196	310
Taux (%)	7,10	7,10	22,59	63,23	100

Au niveau des vrais positifs (63,23% des échantillons), une régression linéaire est réalisée entre les valeurs prédites et les valeurs réelles. Les résultats sont présentés dans le Tableau 5 et représenté en pointillées rouges sur la Figure 30.

Tableau 5 : Qualité de modélisation de l'abondance de Pn en cas de présence

Ordonnée à l'origine		Coefficient	
Estimation	Ecart-type	Estimation	Ecart-type
2,16	1,35E-01	4,57E-01	3,28E-02

L'équation de la droite obtenue est loin de l'équation $y=x$ (ordonnée à l'origine égale à 0 et coefficient égal à 1) (Tableau 5). L'ordonnée à l'origine est par ailleurs assez élevée (2,16 au lieu de 0 dans une modélisation plus optimale). La prédiction de l'abondance de Pn est moyenne.

Bilan des modèles réalisés

Les deux modèles obtenus sont **complémentaires**.

Le **premier modèle** a permis de mettre en évidence les facteurs environnementaux impliqués de façon significative dans l'**apparition** de Pn. Une fois celle-ci présente, le **deuxième modèle** met en évidence les facteurs influençant majoritairement sa **croissance** et l'apparition d'**efflorescences**.

Six variables permettent d'expliquer de façon significative l'apparition de Pn dans l'ensemble du Mor Braz d'après le premier modèle : la teneur en **silicates**, la **population phytoplanctonique présente** (sa **diversité** (Hmax), l'**équitabilité de Piélou (J)**) et son **abondance**, représentée par la teneur en chlorophylle a), la **température** et dans une moindre mesure le **débit de la Vilaine**.

D'après le deuxième modèle, **cinq** variables environnementales ont une influence significative sur l'abondance de Pn quand elle est présente : la **stratification moyenne**, la teneur en **silicates** et le rapport **phosphates/silicates** dans l'eau, l'**énergie utile moyenne** reçue, l'**équitabilité de Piélou (J)**.

V. Discussion et conclusion

L'**objectif** de ce stage était d'analyser les **relations** statistiques entre les pics d'abondance appelées **efflorescences de *P. nitzschia*** et les **conditions environnementales dans le Mor Braz**. Afin d'atteindre cet objectif, les données d'abondance de *P.nitzschia* acquises par le RÉPHY au niveau de deux sites géographiquement éloignés dans le Mor Braz (Ouest Loscolo et Men er Roue) ont été analysées.

Différentes méthodes statistiques ont été utilisées pendant ce stage avant de choisir la modélisation de type **GAM**. En effet, les séries temporelles des données analysées dans le cadre de ce stage n'ont pas une fréquence régulière et elles comportent de nombreuses données manquantes. La loi de distribution de l'abondance de *P.nitzschia* est « particulière ». Ces deux constats m'ont poussée à rejeter l'utilisation de l'analyse temporelle ou la méthode « zero-inflated ». De plus les relations entre l'abondance de *P.nitzschia* et les nombreuses variables environnementales ne sont pas linéaires ainsi que le logit de sa probabilité de présence. La modélisation de type GAM a donc été retenue. Ce type de méthode, qui m'avait été fortement conseillé, avait déjà été utilisé pour traiter des données semblables d'abondance de phytoplancton (Zuur *et al.*, 2009)..

Le GAM est un modèle **non paramétrique**. L'avantage a été de pouvoir modéliser les relations non linéaires. Toutefois, si les relations entre l'abondance de *P.nitzschia* et les conditions environnementales ont pu être mises en évidence, elles n'ont pas pu être quantifiées (pas de coefficients, pas d'Odds ratio). D'autres méthodes de sélection de variables (selon le GCV score/ l'UBRE (proche de l'AIC)...) existent. Il serait intéressant de vérifier que les résultats obtenus avec ces méthodes concordent avec ce qui a été obtenu durant ce stage et de tenter de quantifier ces influences.

L'approche proposée dans le rapport est de décomposer l'abondance de *P.nitzschia* observée en deux phénomènes complémentaires : l'**apparition** et la **croissance**/multiplication. Ces deux phénomènes ont été étudiés séparément (deux modèles construits) puis les résultats ont été rassemblés (combinaison des deux modèles).

Les résultats descriptifs des données ont montré le caractère **saisonnier** de *P.nitzschia* dans le Mor Braz. Les efflorescences de *P.nitzschia* apparaissent généralement au **printemps/été** et leur intensité, fréquence et dates d'apparition diffèrent entre les points de suivi Ouest Loscolo et Men er Roue. D'autres efflorescences, plus tardives dans la saison, apparaissent également en **automne** à **Ouest Loscolo**, contrairement à ce qui a été observé à Men er Roue. Ces résultats sont en accord avec ceux de la bibliographie qui décrit *P.nitzschia* comme une espèce connue pour ses efflorescences saisonnières qui peuvent durer jusqu'à **deux mois** (Bates *et al.*, 1989 ; Martin *et al.*, 1990 ; Fryxell *et al.*, 1997 ; Gallacher *et al.*, 2001),

Les périodes de printemps/été sont des périodes où de nombreux phytoplanctons apparaissent. La majorité de ce phytoplancton est constituée de diatomées, dont fait parti *P.nitzschia*. La population phytoplanctonique est plus **abondante, diversifiée et déséquilibrée** en printemps/été confirmant l'apparition et la croissance de nombreuses espèces. Ce déséquilibre est représenté par **l'équitabilité de Piélou**, sélectionné dans les deux modèles.

La **teneur en silicates** et **l'équilibre de la population phytoplanctonique** ont un rôle significatif pour **l'apparition** et la **croissance** de *P.nitzschia*. Ces deux variables semblent liées au caractère compétitif de *P.nitzschia*. Sa compétitivité face aux autres phytoplanctons a été supposée liée à sa capacité à croître lorsque la teneur en nutriments est limitée (Trainer *et al.*, 2012 ; Downes-Tettmar *et al.*, 2013) mais aussi à la production de l'acide domoïque (toxine ASP responsable des intoxications) (MacIntyre *et al.*, 2011). La production d'acide domoïque par *P.nitzschia* a été significativement mise en corrélation avec la faible teneur en silicates (Anderson *et al.*, 2006 et MacIntyre *et al.*, 2011). Les conditions de croissance et de multiplication de *P.nitzschia* ne sont en effet pas les mêmes que celles de production d'acide domoïque (Thessen *et al.*, 2009 ; Lelong *et al.*, 2012). De plus le genre *P.nitzschia* comporte plusieurs espèces de toxicité différente, *P.nitzschia australis* étant soupçonnée comme étant en majorité l'origine de la contamination des coquillages. La poursuite de l'étude pourrait donc se concentrer sur chaque espèce précise de Pn, leur **comportement face à la limitation des nutriments** (en particulier les silicates) et la production d'acide domoïque.

L'apparition de fortes efflorescences au dessus du seuil d'alerte de 100 000 à 300 000 cellules/L et mises en cause dans la contamination des coquillages et les intoxications par l'ASP n'a pas été expliquée au cours de ce stage. La poursuite de l'étude pourrait se concentrer sur chaque

espèce précise de *P.nitzschia*, leur **comportement face à la limitation des nutriments** (en particulier les silicates) et la production d'acide domoïque.

Si la **température** est significative pour expliquer les dates d'**apparition** de *P.nitzschia*, **l'énergie utile moyenne** a un rôle significatif pour sa multiplication et sa **croissance**. La température est fortement corrélée aux saisons. La significativité de cette variable exprime essentiellement le caractère saisonnier de *P.nitzschia* dans le Mor Braz. Une étude effectuée aux Etats-Unis a montré que les efflorescences de *P.nitzschia* apparaissent sur un large ensemble de température (de 10 à 30 °C) (Andrew J. Shuler *et al.*, 2012). Cette étude a supposé que le rayonnement était plus influent que la température pour expliquer la croissance et les efflorescences de *P.nitzschia*. Ces résultats sont en accord avec ceux obtenus avec le modèle de croissance développé au cours du stage. Ce n'est en effet pas la température mais l'énergie utile moyenne, c'est-à-dire le rayonnement réellement reçu (la semaine précédente) qui est associée à la croissance de *P.nitzschia* une fois qu'il est apparu.

La **stratification moyenne** de l'eau de mer joue un rôle très significatif dans la croissance de *P.nitzschia* dans le Mor Braz, que ce soit à Ouest Loscolo ou à Men er Roue. Cette stratification est elle-même liée au **vent**. Lorsque le vent augmente, l'eau de mer est remuée et la stratification diminue. Dans le Golfe du Mexique, les efflorescences de *P.nitzschia* apparaissent en automne lorsque les vents diminuent la stratification de l'eau et en été lors des grands débits des fleuves et d'autres apparaissent (Dortch *et al.*, 1997). Dans notre étude, le débit de la Vilaine n'apporte que peu d'informations.

La corrélation avec la **salinité** ou les macronutriments/**nitrites** (très corrélés) n'a pas été mise en évidence durant ce stage. Les efflorescences de *P.nitzschia* apparaissent sur un ensemble large de salinité (de 15 à 37), mais ceci est lié à la présence de différentes espèces au sein du même genre *P.nitzschia* dont les relations avec la salinité diffèrent (Andrew J. Shuler *et al.*, 2012 ; Downes-Tettmar *et al.*, 2013). Ces études montrent, en outre, que selon les espèces, *P.nitzschia* peut être influencée de façon significative par les phosphates, la salinité, les macronutriments ou les **nitrites**.

Les trois nutriments (azote, phosphates et silicates) sont essentiels pour le développement de *P.nitzschia*. Le rapport **phosphates/silicates** (P/Si) a été mis en évidence pour expliquer la croissance de *P.nitzschia* dans le Mor Braz. Le silicate est le nutriment ayant un **rôle décisif** dans le Mor Braz d'après les résultats obtenus.

La teneur en **silicates** a été sélectionnée par les deux modèles. Ce nutriment influence donc l'apparition et la croissance de *P.nitzschia*. Étant une diatomée, elle possède un squelette construit à partir de silice et les silicates sont une source exploitable et essentielle pour elle. Or la croissance et l'apparition de *P.nitzschia* sont majoritairement corrélées à la **baisse de la teneur en silicates**. Des

corrélations négatives entre l'abondance et les nutriments ont été démontrées également dans d'autres études (Lelong *et al.*, 2012 ; Downes-Tettmar *et al.*, 2013) confirmant nos résultats. Les silicates étant cependant consommés par *P.nitzschia* et autres diatomées, il est difficile d'établir si la limitation en silicates est un facteur ou un résultat de l'apparition des efflorescences de *P.nitzschia* dans le Mor Braz.

Les données du RÉPHY, dans le Mor Braz, sont des **données de surveillance**. Aucun plan d'analyses statistiques n'a été réalisé avant l'acquisition de ces données. L'échantillonnage est irrégulier. Les prélèvements ne sont pas effectués sur les deux sites Ouest Loscolo et Men er Roue à la même date. Les heures, les méthodes et les niveaux de prélèvements changent au cours du temps. Beaucoup de données sont manquantes, notamment en ce qui concerne les nutriments mais également les données hors RÉPHY (périodes disponibles restreintes). Plusieurs problèmes ont donc été rencontrés pour traiter ces données. Ce sont autant de facteurs augmentant la variabilité dans les données et l'**incertitude** dans les résultats.

La thermocline a été observée comme étant le niveau de profondeur où le phytoplancton est le plus abondant (compromis entre lumière et teneur en nutriments) (Longhi *et al.*, 2009). Il serait intéressant de vérifier si l'abondance du phytoplancton mesurée à Ouest Loscolo et Men er Roue dépend significativement du **niveau de profondeur**. Pour cela, il serait nécessaire d'acquérir des données de plusieurs niveaux de profondeur à la même date (données absentes actuellement).

D'autres variables ont été étudiées et mises en corrélation avec la croissance de *P.nitzschia* (métaux, urée, glutamate...) (Thessen *et al.*, 2009 ; Lelong *et al.*, 2012) mais ne sont pas disponibles donc pas analysées dans notre étude.

P.nitzschia est consommée par de nombreux organismes filtreurs de phytoplancton (des dinoflagellés hétérotrophes aux poissons). Cela n'a pu être pris en compte dans ce stage bien que l'abondance de *P.nitzschia* en soit affectée. La **consommation** pourrait par ailleurs expliquer la fin de l'efflorescence.

Notre étude a ciblé le genre *P.nitzschia*. Cependant la **variabilité inter-espèce** est grande, chaque espèce ayant un comportement différent face aux conditions environnementales (Thessen *et al.*, 2009 ; Lelong *et al.*, 2012 ; Trainer *et al.*, 2012). La **variabilité intra-espèce** est également importante (Lelong *et al.*, 2012). Analyser l'ensemble des espèces rassemblées au genre *P.nitzschia* augmente donc la difficulté de mettre en évidence l'effet d'une variable environnementale, mais l'identification des cellules de phytoplancton par le RÉPHY ne permet pas actuellement d'étudier chaque espèce de *P.nitzschia* présente dans le Mor Braz.

L'ordre est un niveau taxonomique assez haut dans la classification. La qualité d'identification du phytoplancton a poussé à choisir ce niveau taxonomique. Cependant on peut se demander si les **indices écologiques** basés sur l'ordre sont un bon descripteur de la population phytoplanctonique.

Les données de l'étude peuvent être analysées selon des **approches différentes**. L'approche proposée dans le rapport est de décomposer l'abondance de *P.nitzschia* observée en deux phénomènes complémentaires : l'apparition et la croissance. L'effet des variables n'est pas significativement différent entre Ouest Loscolo et Men er Roue. La moyenne de l'abondance de *P.nitzschia* est cependant supposée constante dans les modèles construits. Cette moyenne est sans doute différente : cela expliquerait peut-être le décalage observé lors de la validation croisée.

Une autre approche a été réalisée en posant la moyenne de l'abondance de *P.nitzschia* comme étant dépendante du temps. Cette moyenne, appelée moyenne saisonnière, et les résidus restant (différence entre l'abondance de *P.nitzschia* et la moyenne saisonnière) ont été extraits et étudiés séparément entre Ouest Loscolo et Men er Roue. Cette autre approche amène d'autres résultats qu'il serait intéressant de vérifier. La moyenne saisonnière est notamment mise en relation linéaire avec la teneur en silicate.

Que ce soit avec l'une ou l'autre approche, les fortes efflorescences de *P.nitzschia*, au dessus du seuil d'alerte de 100 000 à 300 000 cellules/L, restent difficiles à expliquer. Les modèles obtenus ne permettent pas de les prédire. Il manque peut-être une information pour y remédier. De plus, les relations mises en évidence avec le GAM sont difficiles à interpréter et de grandes incertitudes persistent (données manquantes, valeurs extrêmes peu observées dans le milieu...).

En conclusion, ce stage a permis d'atteindre l'objectif principal, en mettant en évidence les conditions environnementales qui influencent significativement l'abondance et les efflorescences de *P.nitzschia* dans le Mor Braz. Ce sont les silicates et le rapport phosphates/silicates, la stratification moyenne, la température ou l'énergie utile moyenne et les caractéristiques de la communauté phytoplanctonique présente.

D'une façon secondaire, le stage a permis aussi d'analyser la **qualité** de certaines données acquises par Ifremer concernant le phytoplancton et de les **valoriser**, d'étudier le phytoplancton dans son ensemble au niveau du Mor Braz (communauté phytoplanctonique présente, évolution dans le

temps...) et enfin d'étudier globalement les conditions environnementales dans le Mor Braz (richesse en nutriments, écoulements des fleuves...).

Ce stage alliait **santé et environnement** dans une même problématique. Il m'a permis ainsi d'utiliser d'une part mes connaissances en biologie et environnement et d'autre part celles acquises cette année en statistique et informatique pour la santé. Mes compétences relatives à l'utilisation du logiciel *R* et son interface *RStudio* ont été grandement améliorées. Le GAM est une technique de modélisation qui a été étudiée et utilisée au cours de ce stage, me permettant de travailler avec des données environnementales, différentes de celles vues en cours. Travailler au sein d'une équipe et rencontrer des professionnels ou chercheurs m'a apporté des compétences en gestion du temps et communication (réalisation de présentations, transfert et échange de connaissances, compte-rendus...).

VI. Liste des abréviations

- **Pn** : *Pseudo-nitzschia*
- **IFREMER** : Institut Français de Recherche pour l'Exploitation de la MER
- **RÉPHY** : Réseau national de surveillance du phytoplancton et des phycotoxines dans les coquillages
- **OL** : Ouest Loscolo
- **MeR** : Men er Roue
- **Abondance** : Ce qui est nommé « Abondance » dans le rapport correspond à la concentration en cellules de phytoplancton par litre d'eau de mer.
- **N** : Azote inorganique
- **SiOH** : Silicates
- **PO4** : Phosphates
- **Mies** : Matière inorganique en suspension
- **Chlo a** : Chlorophylle a
- **GAM** : Modèle Additif Généralisé
- **NASA** : National Aeronautics and Space Administration

VII. Table des illustrations

VII.1. Table des tableaux

Tableau 1 : Variables explicatives et coefficients associés du modèle de GAM logistique obtenu	30
Tableau 2 : Matrice de confusion du modèle en prenant un seuil de 0,5	31
Tableau 3 : Le GAM expliquant la croissance de Pn	33
Tableau 4 : Prédiction de l'absence et de la présence par le modèle combiné	36
Tableau 5 : Qualité de modélisation de l'abondance de Pn en cas de présence	36
Tableau 6 : Variables, unités et méthodes de mesure du RÉPHY selon les périodes	53
Tableau 7 : Période disponibles pour les données extérieures.....	54
Tableau 8 : Taxons et fréquences associés au genre <i>P.nitzschia</i>	56
Tableau 9 : Exemple de classification des taxons	57
Tableau 10 : Matrice des corrélations de Spearman entre les variables environnementales utilisées	59
Tableau 11 : Paramètres de la régression linéaire expliquant la moyenne saisonnière de l'abondance de Pn à OL.....	68
Tableau 12 : Paramètres de la régression linéaire expliquant la moyenne saisonnière de l'abondance de Pn à MeR	69
Tableau 13 : GAM construit pour expliquer les fluctuations de l'abondance de Pn autour de sa moyenne saisonnière à OL	71
Tableau 14 : GAM construit pour expliquer les fluctuations de l'abondance de Pn autour de sa moyenne saisonnière à MeR	72
Tableau 15 : Abondance de Pn mise en classes	78
Tableau 16 : Corrélations entre les dimensions de l'ACP et les variables quantitatives	81
Tableau 17 : Variable qualitatives expliquées par l'ACP	81
Tableau 18 : Positionnement des catégories des variables qualitatives par rapport aux axes de l'ACP	82
Tableau 19 : Classes obtenues ordonnées par rapport aux moyennes de chaque variable pour chaque classe (des plus forte aux plus faibles moyennes).....	84

VII.2. Table des figures

Figure 1 : Carte du Mor Braz et localisation des points de suivi du RÉPHY Ouest Loscolo et Men er Roue	8
Figure 2 : Prélèvement d'eau en mer et comptage des cellules de phytoplancton	9
Figure 3 : Exemple de fonctionnement de l'analyse des efflorescences de Pn (données de MeR en 2006)	14
Figure 4 : Schéma de l'approche multivariée expliquant l'abondance de Pn	16
Figure 5 : Exemple de mise en forme des données du RÉPHY	19
Figure 6 : Construction de la base de données exploitée	23
Figure 7 : L'abondance de Pn à OL de 2000 à 2005	24
Figure 8 : Saisonnalité de l'abondance de Pn à Ouest Loscolo	25
Figure 9 : Saisonnalité de l'abondance de Pn à Men er Roue	25
Figure 10 : Tendence de l'abondance de Pn à OL.....	25

Figure 11 : Tendance de l'abondance de Pn à MeR.....	25
Figure 12 : Fréquence des efflorescences en fonction du quart de mois de l'année à OL	26
Figure 13 : Fréquence des efflorescences en fonction du quart de mois de l'année à MeR	26
Figure 14 : Intensité et dates des efflorescences de Pn le long de la période d'étude à OL.....	26
Figure 15 : Intensité et dates des efflorescences de Pn le long de la période d'étude à MeR	26
Figure 16 : Fréquence de la durée des efflorescences de Pn en nombre de jours à OL	27
Figure 17 : Fréquence de la durée des efflorescences de Pn en nombre de jours à MeR	27
Figure 18 : L'abondance de Pn en fonction de la teneur en SiOH dans l'eau de mer	28
Figure 19 : L'abondance de Pn en fonction de l'énergie utile moyenne	28
Figure 20 : Histogramme de l'abondance de PN sur l'ensemble des données	29
Figure 21 : Probabilité de présence de Pn données par le GAM logistique en fonction de la présence ou absence de Pn réelle	31
Figure 22 : Courbe ROC du GAM logistique obtenu	31
Figure 23 : Résidus de Pearson des données modélisés par le GAM logistique	32
Figure 24 : Courbe ROC du GAM logistique lors de la validation croisée	32
Figure 25 : Prédiction du GAM lors de la validation croisé	32
Figure 26 : Résidus de Pearson du GAM obtenu	34
Figure 27 : Abondance de Pn prédite par le modèle en fonction de l'abondance de Pn réelle.....	34
Figure 28 : Prédiction du GAM sur les données en présence de Pn suite à la validation croisée	35
Figure 29 : Prédiction du deuxième modèle sur l'ensemble des données	35
Figure 30 : Prédiction du deuxième modèle une fois combiné au premier modèle	36
Figure 31 : Exemple d'interface d'extraction de Quadrigé ²	52
Figure 32 : Le principe du lissage par régression cubique	55
Figure 33 : Principe du lissage par régression cubique pénalisée	55
Figure 34 : Pourcentage de cellules non identifiées à l'ordre à OL ou MeR	57
Figure 35 : Boxplot par mois de l'abondance de Pn	58
Figure 36 : Boxplot par année de l'abondance de Pn	58
Figure 37 : Logit observé pour le rayonnement moyen.....	60
Figure 38 : Logit observé pour la température (temp)	60
Figure 39 : Logit observé pour la teneur en silicates (SiOH)	60
Figure 40 : Logit observé pour la teneur en chlorophylle a (chlo a)	60
Figure 41 : Logit observé pour la richesse.....	61
Figure 42 : Logit observé pour l'équitabilité de Piélou	61
Figure 43 : Logit ajusté pour la teneur en SiOH.....	62
Figure 44 : Logit ajusté pour la température	62
Figure 45 : Logit ajusté pour la richesse.....	62
Figure 46 : Logit ajusté pour la teneur en chlo a	62
Figure 47 : Logit ajusté pour l'équitabilité de Piélou	62
Figure 48 : Logit ajusté pour le débit de la Vilaine	62
Figure 49 : Normalité des résidus du GAM expliquant la croissance de Pn	63
Figure 50 : Relation ajustée avec la teneur en SiOH	63
Figure 51 : Relation ajustée avec l'énergie utile moyenne (raycoext_moy)	63
Figure 52 : Relation ajustée avec P/Si.....	63
Figure 53 : Relation ajustée avec J.....	63
Figure 54 : Relation ajustée avec la stratification moyenne (gradnrj_moy)	64

Figure 55 : Saisonnalité de la teneur en SiOH dans l'eau de mer à OL	64
Figure 56 : Saisonnalité de la teneur en SiOH dans l'eau de mer à MeR	64
Figure 57 : Tendances de la teneur en SiOH dans l'eau de mer à OL	65
Figure 58 : Tendances de la teneur en SiOH dans l'eau de mer à MeR	65
Figure 59 : Saisonnalité de la teneur en chlo a dans l'eau de mer à OL	65
Figure 60 : Saisonnalité de la teneur en chlo a dans l'eau de mer à MeR	65
Figure 61 : Tendances de la teneur en chlo a de l'eau de mer à OL	66
Figure 62 : Tendances de la teneur en chlo a de l'eau de mer à MeR	66
Figure 63 : Décomposition de l'abondance de Pn à OL en 2005	67
Figure 64 : Décomposition de l'abondance de Pn à MeR en 2005	67
Figure 65 : Moyennes saisonnières de la teneur en SiOH et de l'abondance de Pn à OL en 2005	67
Figure 66 : Moyennes saisonnières du rayonnement journalier et de l'abondance de Pn à MeR en 2005	67
Figure 67 : Valeurs prédites par le modèle linéaire de la moyenne saisonnière de l'abondance de Pn à OL en fonction de ses valeurs réelles	69
Figure 68 : Valeurs prédites par le modèle linéaire de la moyenne saisonnière de l'abondance de Pn à MeR en fonction de ses valeurs réelles	70
Figure 69 : Validation du modèle explicatif de la moyenne saisonnière de l'abondance de Pn à OL ...	70
Figure 70 : Validation du modèle explicatif de la moyenne saisonnière de l'abondance de Pn à MeR	71
Figure 71 : Valeurs prédites des fluctuations autour de la moyenne saisonnière de l'abondance de Pn à OL en fonction des valeurs réelles	72
Figure 72 : Valeurs prédites des fluctuations autour de la moyenne saisonnière de l'abondance de Pn à MeR en fonction des valeurs réelles	73
Figure 73 : Validation du GAM expliquant autour de la moyenne saisonnière de l'abondance de Pn à OL	73
Figure 74 : Validation du GAM expliquant autour de la moyenne saisonnière de l'abondance de Pn à MeR	74
Figure 75 : Projection des variables actives sur les dimensions 1 et 2 de l'ACP	75
Figure 76 : Position des modalités des variables supplémentaires qualitatives sur les axes 1 et 2 de l'ACP	75
Figure 77 : Projection des variables actives sur les dimensions 3 et 4	76
Figure 78 : Position des modalités des variables supplémentaires qualitatives sur les axes 3 et 4 de l'ACP	76
Figure 79 : Projection des variables actives sur les dimensions 1 et 2	77
Figure 80 : Position des modalités des variables supplémentaires qualitatives sur les axes 1 et 2 de l'ACP	77
Figure 81 : Vecteurs des variables environnementales par rapport aux 2 premiers axes de l'ACP	78
Figure 82 : Positionnement des modalités des variables qualitatives par rapport aux dimensions 1 et 2 de l'ACP	79
Figure 83 : Inertie donnée par les dimensions de l'ACP et dendrogramme de la classification effectuée	83
Figure 84 : Représentations des classes sur le plan des dimensions 1 et 2 de l'ACP	83

VIII. Bibliographie

ANDERSON C.R., BRZEZINSKI M.A., WASHBURN L., KUDELA R., «Circulation and environmental conditions during a toxigenic *Pseudo-nitzschia australis* bloom in Santa Barbara Channel, California», *Marine Ecology Progress Series* n°327, 2006, pp. 119–133.

BATES S.S., BIRD C.J.W., DE FREITAS A.S., FOXALL R., GILGAN M., HANIC L.A., JOHNSON G.R., McCULLOH A.W., ODENSE P., POCKLINGTON R., QUILLIAM M.A., SIM P.G., SMITH J.C., SUBBA RAO D.V., TODD E.C.D., WALTER J.A., WRIGHT J.L.C., «Pennate diatom *Nitzschia pungens* as the primary source of domoic acid, a toxin in shellfish from eastern prince Edward island, Canada», *Canadian Journal of Fisheries and Aquatic Science* n°46, 1989, pp. 1203–1215.

BATES S.S., GARRISON D.L., HORNER R.A., «Bloom dynamique and phyecology of domoic-acid-producing *Pseudo-nitzschia* species», dans : ANDERSON D.M., CEMBELLA A.D., HALLEGRAEFF G.M., *Physiological Ecology of Harmful Algal Blooms*, 1998, pp. 267–292.

CASTELEYN G., CHEPURNOV V.A., LELIAERT F., MANN D.G., BATES S.S., LUNDHOLM N., RHODES L., SABBE K., VYVERMAN W., «*Pseudo-nitzschia pungens* (Bacillariophyceae): a cosmopolitan diatom species? », *Harmful Algae* n°7, 2008, pp. 241–257.

COWPERTWAIT P.S.P., METCALFE A.V., *Introductory time series with R*, Springer, Baltimore (EU), 2009, 254 p.

CRAWLEY M.J., *The R book*, Wiley, West Sussex (England), 2007, 942 p.

DORTCH Q., ROBICHAUX R., POOL S., MILSTED D., MIRE G., RABALAIS N.N., SONIAT T.M., FRYXELL G.A., TURNER R.E., PARSONS M.L., «Abundance and vertical flux of *Pseudo-nitzschia* in the northern Gulf of Mexico», *Marine Ecology Progress Series* n°146, 1997, pp. 249–264.

DOWNES-TETTMAR N., ROWLAND S., WIDDICOMBE C., WOODWARD M., LLEWELLYN C., «Seasonal variation in *Pseudo-nitzschia* spp. and domoic acid in the Western English Channel», *Continental Shelf Research* n°53, 2013, pp. 40–49.

FRYXELL G.A., VILLAC M.C., SHAPIRO L.P., «The occurrence of the toxic diatom genus *Pseudo-nitzschia* (Bacillariophyceae) on the West Coast of the USA, 1920–1996: a review», *Phycologia* n°36, 1997, pp. 419–437.

GALLACHER S., HOWARD G., HESS P., McDONALD E., KELLY M.C., BATES L.A., BROWN N., MCKENZIE M., GILLIBRAND P., TURELL W.R., «The occurrence of amnesic shellfish poisons in shellfish from Scottish waters», Dans : HALLEGRAEFF G.M., BLACKBURN S.I., BOLCH C.J., LEWIS D., «Proceedings of the Ninth International Conference on Harmful Algal Blooms», *Harmful Algal Blooms*, 2000, pp. 30–33.

HASLE G.R., «Are most of the domoic acid-producing species of the diatom genus *Pseudo-nitzschia* cosmopolites ? », *Harmful Algae* n°1, 2002, pp. 137–146.

HICKEY B.M., TRAINER V.L., KORSO P.M., ADAMS N.G., CONOLLY T.P., KACHEL N.B., GEIR S.L., «A springtime source of toxic *Pseudo-nitzschia* cells on razor clam beaches in the Pacific Northwest», *Harmful Algae* n°25, 2013, pp. 1-14.

KLEIN C., CLAQUIN P., BOUCHART V., LE ROY B., VÉRON B., «Dynamics of *Pseudo-nitzschia* spp. and domoic acid production in a macrotidal ecosystem of the Eastern English Channel (Normandy, France) », *Harmful Algae* n°9, 2010, pp. 218–226.

LELONG A., HÉGARET H., SOUDANT P., «Cell-based measurements to assess physiological status of *Pseudo-nitzschia* multiseriis, a toxic diatom», *Research in Microbiology* n° 162, 2011, pp. 969-981.

LELONG A., HÉGARET H., SOUDANT P., BATES S.S., «*Pseudo-nitzschia* (Bacillariophyceae) species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms», *Phycologia* n°51, 2012, pp. 168-216.

LELONG A., JOLLEY D., HÉGARET H., SOUDANT P., «Impact of copper exposure on *Pseudo-nitzschia* spp. physiology and domoic acid production», *Aquatic Toxicology* n°118– 119, 2012, pp. 37– 47.

LEWIS N.I., BATES S.S., McLACHLAN J.L., SMITH J.C., «Temperature effects on growth, domoic acid production, and morphology of the diatom *Nitzschia pungens* f. *multiseriis*», *Elsevier Science Publishers B.V.*, 1993, pp. 601-606.

LONGHI M.L., BEISNER B.E., «Environmental factors controlling the vertical distribution of phytoplankton in lakes», *Journal of Plankton Research* n° 31, 2009, pp. 1195–1207.

MARTIN J.L., HAYA K., BURRIDGE L.E., WILDISH D.J., «*Nitzschia pseudodelicatissima*-a source of domoic acid in the Bay of Fundy, eastern Canada», *Marine Ecology Progress Series* N°67, 1990, pp. 177–182.

SHULER A.J., PATERNOSTER J., BRIM M., NOVOCIN K., TISDALE T., NELLER K., CAHILL J.A., LEIGHFIELD T.A., FIRE S., WANG Z., MORTON S., «Spatial and temporal trends of the toxic diatom *Pseudo-nitzschia* in the Southeastern Atlantic United States», *Harmful Algae* n°17, 2012, pp. 6–13.

TESSON S.V.M, LEGRANDA C., VAN OOSTERHOUT C., MONTRESORA M., KOOISTRAA W.H.C.F, PROCACCINIA G., «Mendelian inheritance pattern and high mutation rates of microsatellite alleles in the diatom *Pseudo-nitzschia multistriata*», *Protist* n°164, 2013, pp. 89–100.

THESSSEN A.E., BOWERS H.A., STOECKER D.K., «Intra- and interspecies differences in growth and toxicity of *Pseudo-nitzschia* while using different nitrogen sources», *Harmful Algae* n°8, 2009, pp. 792-810.

THESSSEN A.E., DORTCH Q., PARSONS M.L., MORISSON W., «Effect of salinity on *Pseudo-nitzschia* species (Bacillariophyceae) growth and distribution», *Journal of Phycology* n°41, 2005, pp. 21–29.

WOOD S.N., *Generalized Additive Models: An introduction with R*, Chapman and Hall/CRC, Boca Raton (EU), 2006, 410 p.

ZUUR A.F., IENO E.N., WALKER N.J., SAVELIEV A.A., SMITH G.M., *Mixed effect models and extensions in ecology with R*, Springer, New-York (EU), 2009, 574 p.

Moteur de recherche : *Google* > <http://www.google.fr/>

Consultation de cours en ligne : Université de Lyon 2 > <http://eric.univ-lyon2.fr/>

Aide à la programmation sur R : L'aide officielle > <http://www.r-project.org/>

Site dédié > <http://www.duclert.org/>

Site dédié > <http://www.statmethods.net/>

Site dédié au package FactoMineR > <http://factominer.free.fr/>

Site dédié au package ggplot 2 > <http://docs.ggplot2.org/>

IX. Annexe

IX.1.	Annexe 1 : L'extraction des données du RÉPHY.....	52
IX.2.	Annexe 2 : Détails sur les données brutes	53
IX.3.	Annexe 3 : Le GAM et les méthodes de lissage.....	55
IX.4.	Annexe 4 : Détails sur la mise en forme de la base	56
IX.4.1.	Les taxons associés au genre <i>P.nitzschia</i>	56
IX.4.2.	Les indices écologiques	56
IX.5.	Annexe 5 : Boxplots de l'abondance de Pn	58
IX.6.	Annexe 6 : Matrice des corrélations de Spearman	59
IX.7.	Annexe 7 : La non-linéarité des logits.....	60
IX.8.	Annexe 8 : Détails des modèles obtenus	61
IX.8.1.	Annexe 8.1 : Le modèle expliquant l'apparition de Pn	61
IX.8.2.	Annexe 8.2 : Le modèle expliquant la croissance de Pn	63
IX.9.	Annexe 9 : Analyses complémentaires	64
IX.9.1.	Annexe 9.1 : Une autre approche pour la modélisation multivariée	64
IX.9.2.	Annexe 9.2 : Analyse de la flore phytoplanctonique	74
IX.9.3.	Annexe 9.3 : Les conditions environnementales rencontrées à OL et MeR.....	77

Indications :

L'Annexe est moins travaillée que le reste du rapport. Les résultats ou discussion présentés dans l'annexe sont incomplets et restent à étudier. Les Annexes 1 à 8 sont directement reliées à ce qui est présenté dans le corps du rapport et en apportent des détails. L'annexe 9 présente des résultats d'analyses complémentaires sur l'abondance de Pn et les conditions environnementales.

IX.1. Annexe 1 : L'extraction des données du RÉPHY

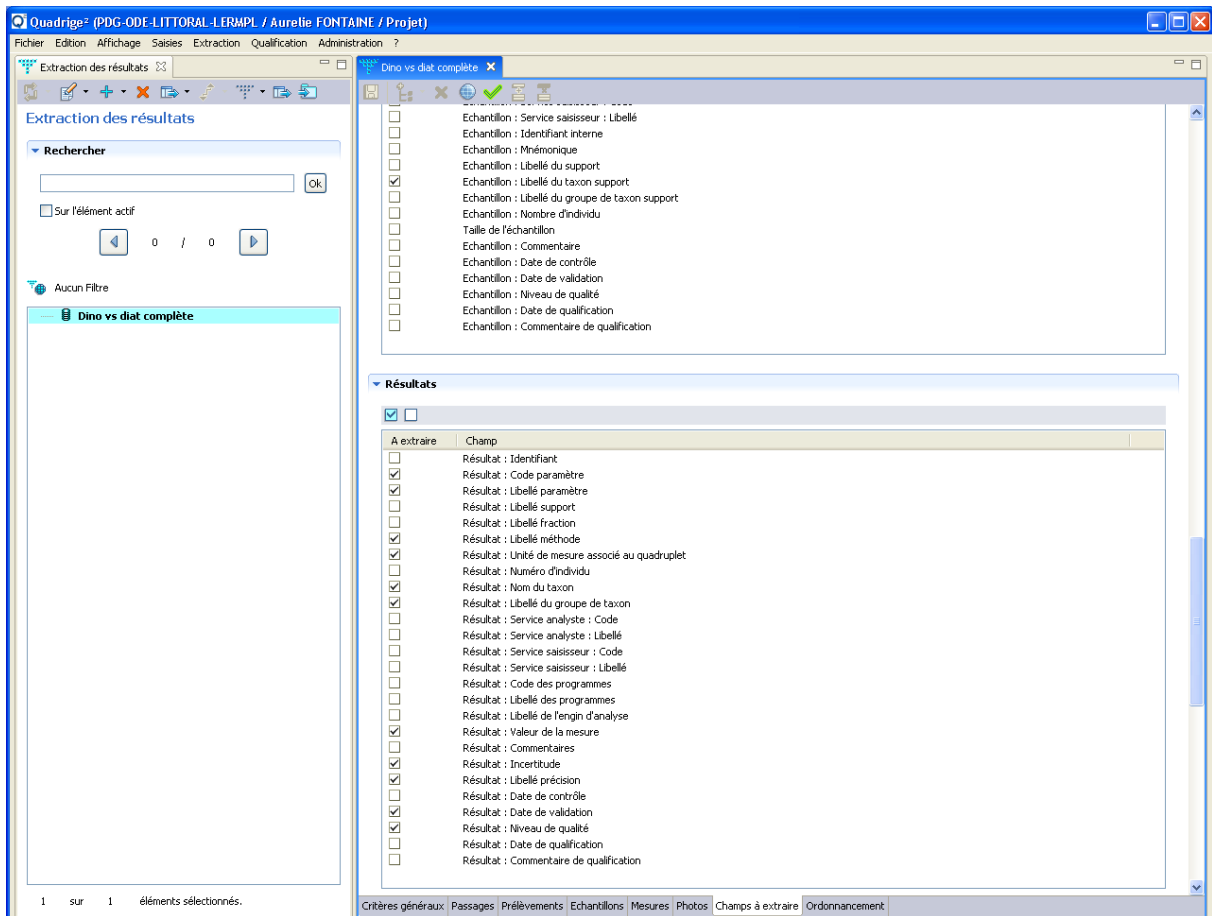


Figure 31 : Exemple d'interface d'extraction de Quadrigé²

IX.2. Annexe 2 : Détails sur les données brutes

Tableau 6 : Variables, unités et méthodes de mesure du RÉPHY selon les périodes

Variable mesurée	Méthode de mesure	Période associée et commentaire	Unité
Température	Capteur de température in situ	de 1984 à 2013	Degré celsius
	Thermomètre à mercure dans échantillon	de 1987 à 2005	
	Thermomètre à renversement	de 1996 à 2005	
	Capteur de température dans bouteille de prélèvement	4 fois, en 2005, validé	
	Thermomètre à mercure in situ	le 29/01/2002 à Men er Roue, non validé	
Salinité	Capteur de conductivité in situ	de 1984 à 2013	sans unité
	Capteur de conductivité dans échantillon	de 1990 à 2010	
	Densimétrie	de 1987 à 1995	
	Titration de Knudsen	4 fois de 1987 à 1995, validé	
Turbidité	Turbidimètre optique (lumière blanche - TURB) dans échantillon	de 1988 à 2007	Nephelometric Turbidity Unit
	Turbidimètre optique (ISO 7027 - TURB FNU) dans échantillon	de 2007 à 2013	Formazine Nephelometric Unit
Ammonium (NH ₄)	Spectrophotométrie manuelle (Aminot et Chaussepied 1983 - Ammonium)	de 1994 à 2007	
	Spectrophotométrie flux (Aminot A. Kérouel R. 2007 - Ammonium)	2 fois en 2007, validé	
	Fluorimétrie flux (Aminot A. Kérouel R. 2007 - Ammonium)	de 2006 à 2012, validé	
Azote nitreux (nitrite NO ₂)	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Nitrite)	de 1994 à 2007	Micromole par litre
	Colorimétrie selon Bend. / Rob. (3). AFNOR	le 29/11/2004 à Ouest Loscolo, validé	
Azote nitrique (nitrate NO ₃)	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Nitrite + nitrate)	de 1994 à 2007	
	Spectrophotométrie manuelle (Aminot et Chaussepied 1983 - Nitrite + nitrate)	le 29/11/2004 à Ouest Loscolo, validé	
Nitrate + nitrite	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Nitrite)	2006	

(NO2+NO3)	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Nitrite + nitrate)	2006 et 2007
	Spectrophotométrie flux (Aminot A. Kérouel R. 2007 - Nitrite + nitrate)	de 2008 à 2012
Silicate (SiOH)	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Silicate)	de 1994 à 2007
	Spectrophotométrie manuelle (Aminot et Chaussepied 1983 - Silicate)	de 1997 à 2007
	Spectrophotométrie flux (Aminot A. Kérouel R. 2007 - Silicate)	de 2008 à 2012
Phosphate (PO4)	Spectrophotométrie flux (Tréguer P., LeCorre P, 1975 - Phosphate)	de 1994 à 2007
	Spectrophotométrie flux (Aminot A. Kérouel R. 2007 - Phosphate)	de 2008 à 2012
	Spectrophotométrie manuelle (Aminot et Chaussepied 1983 - Phosphate)	de 1997 à 2007

Tableau 7 : Période disponibles pour les données extérieures

Variable	Unité	1989	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	2013
Rayonnement	J/cm ²	■																								
Débit de la Loire	m ³ /s	■								■																
Débit de la Vilaine	m ³ /s	■												■												
Chlorophylle a	mg/m ³	■										■														
Spim	mg/m ³	■										■														
Température	° C	■							■																	
Coefficient d'extinction	-	■										■														
Gradient d'énergie	-	■												■												

IX.3. Annexe 3 : Le GAM et les méthodes de lissage

De nombreuses méthodes de lissage peuvent être intégrées au GAM. La méthode de lissage choisie principalement dans ce stage est la régression cubique pénalisée. Il est possible de se référer à la Bibliographie. Son principe est ici expliqué brièvement à partir de la Figure 32.

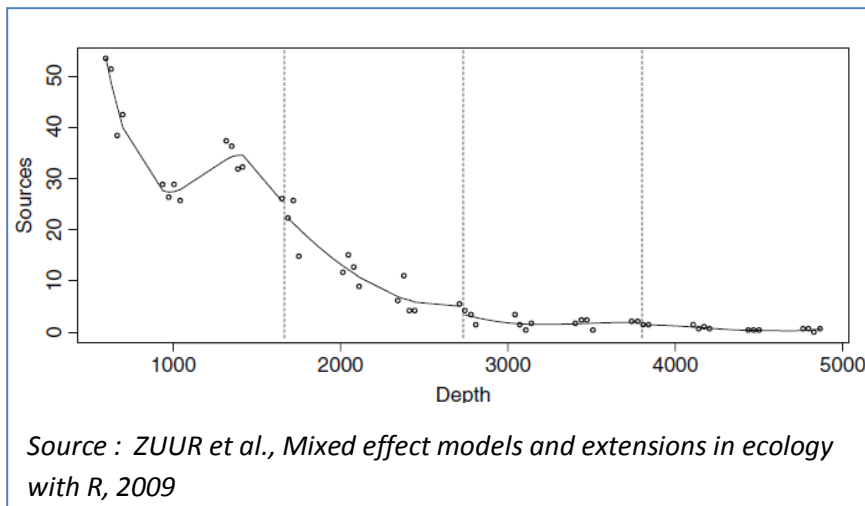


Figure 32 : Le principe du lissage par régression cubique

La variable explicative est découpée en classes. Pour chacune de ces classes une régression cubique est construite en s'ajustant au mieux aux données (de la même façon que pour la régression linéaire). Ensuite des « nœuds » sont créés de façon à relier les différentes courbes obtenues.

Ces régressions sont également pénalisées. Un degré de lissage peut-être spécifié selon le « degré de non linéarité » désiré pour modéliser la relation entre la variable explicative et la variable à expliquer. La Figure 33 illustre différents lissage selon cette pénalisation effectuée.

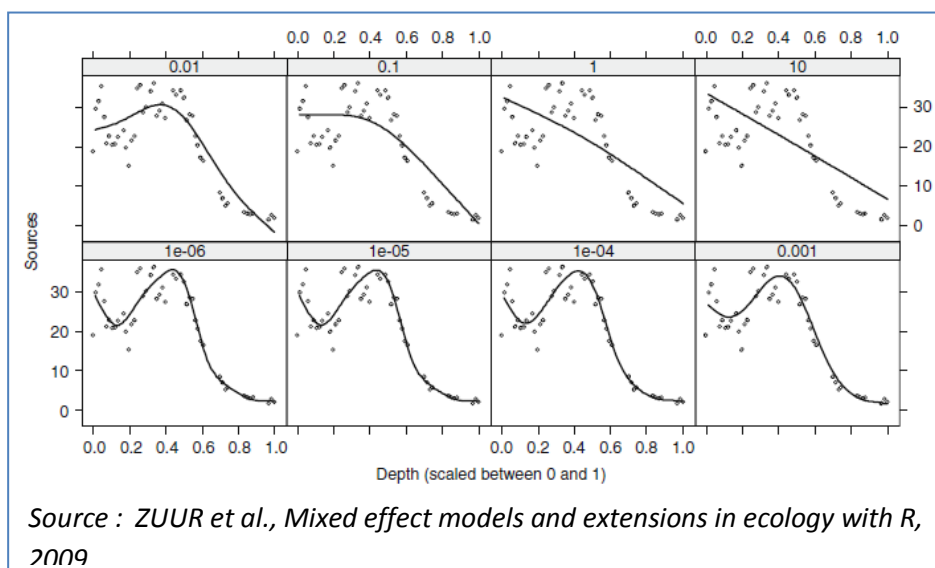


Figure 33 : Principe du lissage par régression cubique pénalisée

IX.4. Annexe 4 : Détails sur la mise en forme de la base

IX.4.1. Les taxons associés au genre *P.nitzschia*

L'abondance de Pn est calculée en additionnant l'abondance de cellules de chaque taxon se référant au genre *P.nitzschia* dans l'échantillon. Onze taxons s'y réfèrent, mais le taxon *Pseudo-nitzschia* (sans distinction de l'espèce) est le plus fréquemment noté (Tableau 8).

Tableau 8 : Taxons et fréquences associés au genre *P.nitzschia*

	Taxon	Effectif (%)
1	Pseudo-nitzschia	589 (70,2)
2	Pseudo-nitzschia, complexe delicatissima, groupe des fines (calliantha + delicatissima + pseudodelicatissima + subcurvata)	83 (9,9)
3	Pseudo-nitzschia, complexe seriata, groupe des effilées (multiseries + pungens)	59 (7,0)
4	Pseudo-nitzschia, complexe seriata, groupe des larges (australis + fraudulenta + seriata + subpacific)	55 (6,5)
5	Pseudo-nitzschia, groupe des sigmoïdes (multistriata)	16 (1,9)
6	Pseudo-nitzschia, groupe des larges symétriques (fraudulenta)	12 (1,4)
7	Pseudo-nitzschia seriata	8 (1,0)
8	Pseudo-nitzschia delicatissima	6 (0,72)
9	Pseudo-nitzschia, complexe americana (americana + brasiliiana)	4 (0,48)
10	Pseudo-nitzschia fraudulenta	4 (0,48)
11	Pseudo-nitzschia, groupe des larges asymétriques (australis + seriata + subpacific)	2 (0,24)

IX.4.2. Les indices écologiques

Les indices écologiques permettent de résumer les informations quant à la **communauté phytoplanctonique présente (sans Pn)** dans chaque échantillon. Ceux utilisés pour l'étude du stage sont les indices J (l'**équitabilité de Piélou**) et Hmax (la **diversité**). Souvent ces indices sont calculés selon l'espèce :

$$J = -\frac{1}{n} \sum_{i=0}^n \ln(p_i) * p_i$$

Hmax=n= Nombre d'espèces différentes rencontrées

p_i= proportion de l'espèce i dans l'échantillon

Cependant, ces indices peuvent être calculés en se basant sur un **autre niveau taxonomique** que l'espèce, même si cela se fait rarement. Le choix de ce niveau taxonomique dépend de l'**identification** des cellules des échantillons prélevés par le RÉPHY.

Tout d’abord, la **classification** de chaque taxon rencontré dans les données extraites de Quadrige² a été effectuée grâce au WORMS. Un extrait de cette classification est présenté dans le Tableau 9. Ce tableau présente trois espèces de phytoplancton toxique rencontrées dans le Mor Braz chaque année. La classification est donnée par un ensemble de niveaux taxonomiques renseignés (ici du super-domaine à l’espèce).

Tableau 9 : Exemple de classification des taxons

Taxon	Alexandrium minutum	Dinophysis acuminata	Pseudo-nitzschia seriata
Super-domaine	Biota	Biota	Biota
Règne	Chromista	Chromista	Chromista
Sous-règne	Harosa	Harosa	Harosa
Infra-règne	Alveolata	Alveolata	Heterokonta
Phylum	Myzozoa	Myzozoa	Ochrophyta
Sous-phylum	Dinozoa	Dinozoa	Khakista
Infra-phylum	Dinoflagellata	Dinoflagellata	
Classe	Peridinea	Peridinea	Bacillariophyceae
Sous-classe	Gonyaulacoidia	Dinophysoidia	Bacillariophycidae
Ordre	Gonyaulacales	Dinophysida	Bacillariales
Famille	Gonyaulacaceae	Dinophysiaceae	Bacillariaceae
Genre	Alexandrium	Dinophysis	Pseudo-nitzschia
Espèce	Alexandrium minutum	Dinophysis acuminata	Pseudo-nitzschia seriata

Souvent l’identification s’arrête au niveau de l’ordre, de la classe, voire même à l’infra-phylum. Par exemple, le technicien note le taxon « Dinoflagellés », ce qui correspond aux « Dinoflagellata », les niveaux taxonomiques inférieures n’étant donc pas renseignés. Il était donc judicieux d’étudier l’identification des cellules phytoplanctoniques par le RÉPHY pour choisir un niveau taxonomique représentant au mieux chaque échantillon de la base utilisée pour le stage (Figure 34).

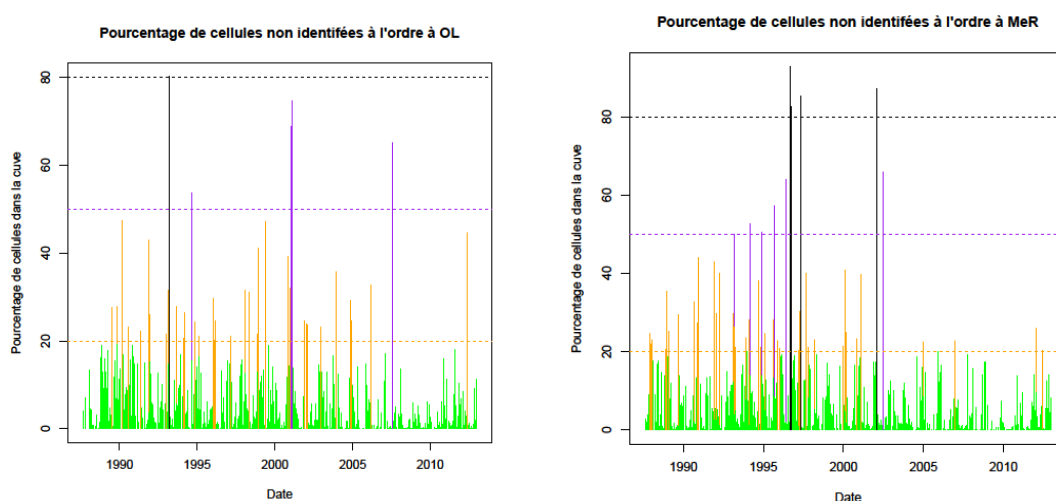


Figure 34 : Pourcentage de cellules non identifiées à l’ordre à OL ou MeR

Les niveaux taxonomiques genre, famille et ordre ont été étudiés. La proportion de cellules non identifiées à ces différents niveaux taxonomiques pour chaque échantillon (c'est-à-dire la proportion en cellules des valeurs manquantes dans la table pour ces différentes colonnes) ont permis de conclure sur le choix de l'ordre pour le calcul des indices écologiques vus précédemment (Figure 34). Pour l'ordre, cette proportion est acceptable (souvent en dessous de 20%, en vert). Au lieu de l'espèce, ces indices sont donc calculés selon l'ordre. L'abondance de Pn est soustraite à celle de l'ordre des Bacillariales dont elle fait parti.

IX.5. Annexe 5 : Boxplots de l'abondance de Pn

L'abondance de Pn et son caractère saisonnier a été visualisée par des boxplots (Figure 35 et Figure 36). Riche en informations, ces boxplots n'ont cependant pas été mis dans le corps du rapport (redondance de résultats similaires).

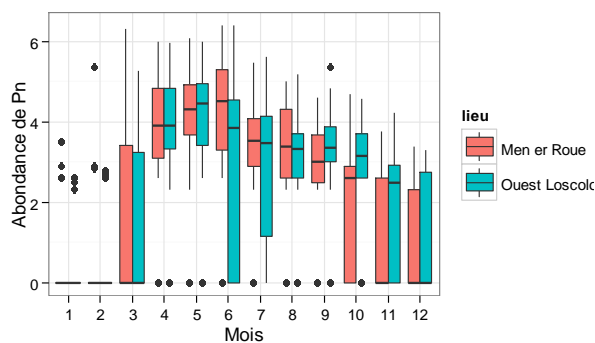


Figure 35 : Boxplot par mois de l'abondance de Pn

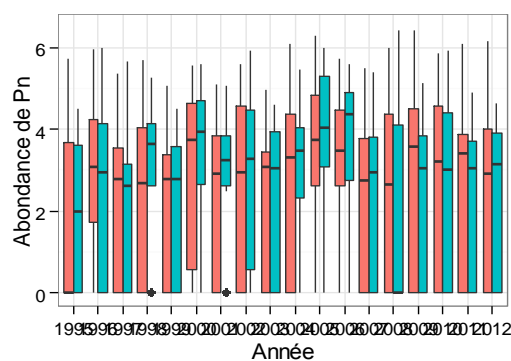


Figure 36 : Boxplot par année de l'abondance de Pn

IX.6. Annexe 6 : Matrice des corrélations de Spearman

La **matrice des corrélations de Spearman** (fournie en annexe) montre de fortes corrélations entre les variables explicatives de l'étude. Cette analyse permet de mettre en garde la sélection des variables dans le modèle. Construire un modèle multivarié avec deux variables explicatives fortement corrélées est à éviter.

Tableau 10 : Matrice des corrélations de Spearman entre les variables environnementales utilisées

	Salin.	Temp.	Turbi.	SiOH	Hmax	J	PO4	Debit de la Loire	Debit de la Vilaine	Chloro. a	Coeff. d'extinct.	Spim	Stratif.	Temp. par sat.	Debit de la Loire moyen	Debit de la Vilaine moyen	Temp. par sat. moy.	Stratif. moy.	Coeff. d'extinct. moyen	N	N/P	N/Si	P/Si	Si/N	Energie utile	Energie utile moy.	Spim moyen	Rayon.
Salinité	1,000																											
Température	0,508	1,000																										
Turbidité	-0,353	-0,362	1,000																									
SiOH	-0,415	-0,453	0,535	1,000																								
Hmax	0,088	0,282	0,065	-0,251	1,000																							
J	-0,029	-0,183	0,133	0,188	-0,188	1,000																						
PO4	-0,392	-0,454	0,640	0,815	-0,207	0,254	1,000																					
Debit de la Loire	-0,630	-0,700	0,344	0,365	-0,251	0,190	0,402	1,000																				
Debit de la Vilaine	-0,699	-0,795	0,423	0,422	-0,309	0,101	0,430	0,888	1,000																			
Chlorophylle a	-0,342	0,238	0,052	-0,158	0,405	-0,265	-0,123	0,005	-0,042	1,000																		
Coefficient d'extinction	-0,505	-0,266	0,679	0,505	0,047	0,083	0,623	0,372	0,388	0,439	1,000																	
Spim	-0,397	-0,365	0,756	0,626	-0,049	0,167	0,721	0,368	0,415	0,063	0,878	1,000																
Stratification	-0,549	-0,107	0,158	0,123	0,181	-0,054	0,182	0,302	0,300	0,492	0,438	0,225	1,000															
Temp. par satellite	0,531	0,922	-0,355	-0,409	0,345	-0,249	-0,446	-0,765	-0,774	0,218	-0,273	-0,367	-0,101	1,000														
Debit de la Loire moyen	-0,648	-0,686	0,308	0,354	-0,243	0,166	0,377	0,985	0,869	0,033	0,349	0,328	0,323	-0,753	1,000													
Debit de la Vilaine moyen	-0,724	-0,735	0,379	0,403	-0,329	0,097	0,389	0,908	0,967	-0,015	0,371	0,377	0,330	-0,771	0,901	1,000												
Temp. par satellite moyenne	0,548	0,911	-0,334	-0,381	0,335	-0,229	-0,417	-0,780	-0,784	0,193	-0,265	-0,340	-0,141	0,990	-0,772	-0,785	1,000											
Stratif. moyenne	-0,599	-0,091	0,206	0,127	0,228	-0,107	0,165	0,286	0,288	0,565	0,484	0,262	0,872	-0,088	0,310	0,317	-0,120	1,000										
Coeff. d'extinct. moyen	-0,533	-0,297	0,671	0,514	0,046	0,073	0,638	0,405	0,413	0,376	0,899	0,807	0,519	-0,304	0,390	0,412	-0,304	0,506	1,000									
N	-0,820	-0,831	0,524	0,558	-0,252	0,183	0,583	0,799	0,817	-0,027	0,512	0,550	0,433	-0,841	0,801	0,823	-0,844	0,402	0,567	1,000								
N/P	-0,717	-0,692	0,212	0,150	-0,157	0,050	0,064	0,695	0,715	0,032	0,213	0,200	0,380	-0,723	0,717	0,747	-0,743	0,348	0,277	0,813	1,000							
N/Si	-0,678	-0,696	0,213	-0,074	-0,110	0,052	0,086	0,692	0,675	0,139	0,254	0,175	0,389	-0,697	0,707	0,703	-0,722	0,359	0,307	0,759	0,886	1,000						
P/Si	0,143	0,096	0,019	-0,472	0,147	0,052	0,054	-0,001	-0,101	0,110	0,087	0,010	0,072	0,007	-0,018	-0,115	-0,001	0,021	0,096	-0,094	-0,134	0,243	1,000					
Si/N	0,678	0,696	-0,213	0,074	0,110	-0,052	-0,086	-0,692	-0,675	-0,139	-0,254	-0,175	-0,389	0,697	-0,707	-0,703	0,722	-0,359	-0,307	-0,759	-0,886	-1,000	-0,243	1,000				
Energie utile	0,351	0,504	-0,631	-0,644	0,180	-0,267	-0,789	-0,374	-0,416	0,035	-0,689	-0,778	-0,079	0,486	-0,332	-0,379	0,448	-0,106	-0,628	-0,571	-0,193	-0,184	-0,042	0,184	1,000			
Energie utile moy.	0,439	0,609	-0,717	-0,711	0,217	-0,286	-0,822	-0,490	-0,538	0,032	-0,706	-0,805	-0,181	0,594	-0,444	-0,503	0,566	-0,162	-0,761	-0,691	-0,307	-0,280	-0,011	0,280	0,871	1,000		
Spim moyen	-0,423	-0,388	0,732	0,623	-0,057	0,172	0,718	0,402	0,429	0,086	0,811	0,896	0,326	-0,404	0,367	0,415	-0,388	0,299	0,902	0,598	0,264	0,235	0,019	-0,235	-0,716	-0,850	1,000	
Rayonnement	0,142	0,479	-0,367	-0,558	0,204	-0,298	-0,627	-0,271	-0,301	0,366	-0,263	-0,449	0,195	0,476	-0,230	-0,257	0,427	0,203	-0,259	-0,441	-0,123	-0,075	0,039	0,074	0,875	0,696	-0,413	1,000
Rayonnement moyen	0,227	0,631	-0,470	-0,657	0,271	-0,367	-0,711	-0,414	-0,459	0,396	-0,320	-0,524	0,183	0,634	-0,359	-0,409	0,591	0,192	-0,328	-0,577	-0,236	-0,159	0,079	0,159	0,767	0,840	-0,518	0,827

Spim est le terme en anglais pour mies : Suspended particular inorganic matter.

IX.7. Annexe 7 : La non-linéarité des logits

Afin de vérifier la linéarité du logit par rapport à chaque variable prédictive du modèle, des procédures graphiques ont été réalisées :

- Chaque variable explicative quantitative est découpée en intervalles (déciles par exemple) ;
- Pour chaque intervalle, la proportion de positifs p est calculée (soit la probabilité de présence) ;
- Le logit observé, soit $\ln(p/(1-p))$, est représenté en fonction de la moyenne de la variable explicative pour chaque intervalle.

Six résultats sont présentés (Figures 36 à 41).

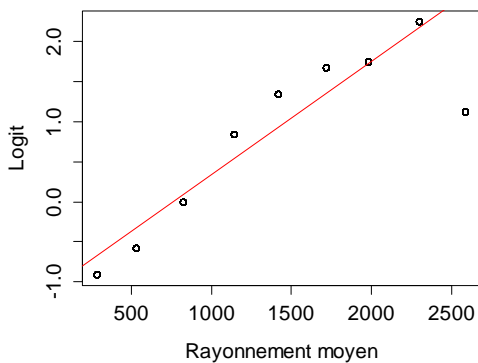


Figure 37 : Logit observé pour le rayonnement moyen

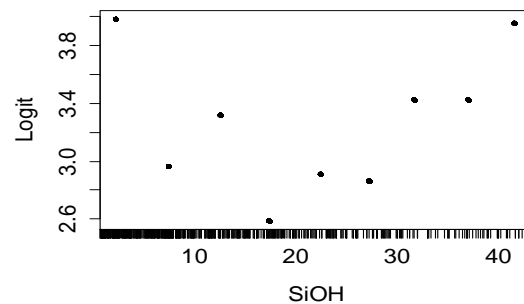


Figure 39 : Logit observé pour la teneur en silicates (SiOH)

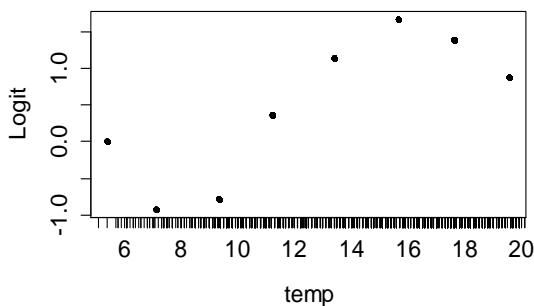


Figure 38 : Logit observé pour la température (temp)

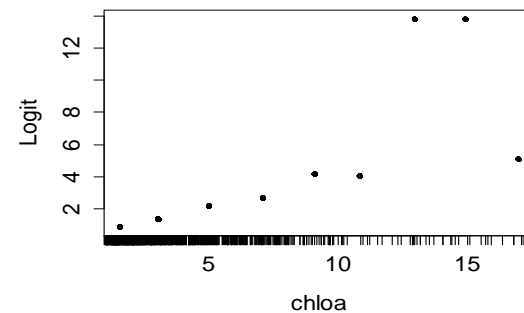


Figure 40 : Logit observé pour la teneur en chlorophylle a (chlo a)

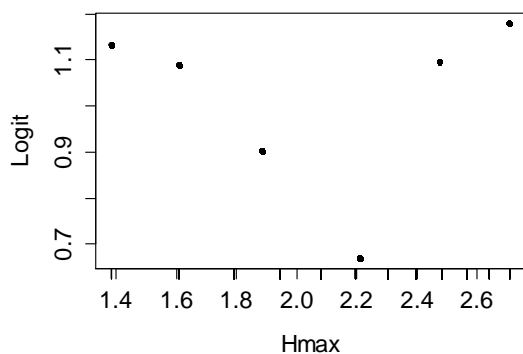


Figure 41 : Logit observé pour la richesse

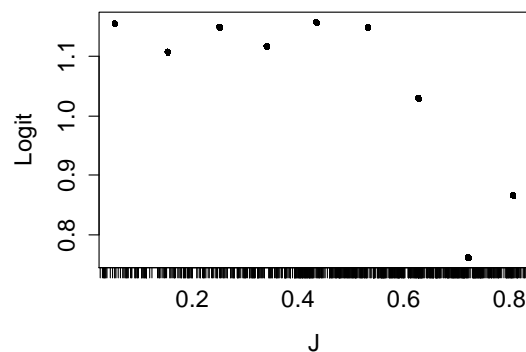


Figure 42 : Logit observé pour l'équitabilité de Piélou

Le logit est plus ou moins linéaire pour le rayonnement moyen (ainsi que pour la chlorophylle a éventuellement, en retirant le dernier point). Cependant il ne peut être considéré comme étant linéaire pour les autres variables. L'hypothèse de **linéarité des logits** est donc **rejetée**. Le GAM logistique est alors utilisé à la place de la régression logistique.

IX.8. Annexe 8 : Détails des modèles obtenus

IX.8.1. Annexe 8.1 : Le modèle expliquant l'apparition de Pn

Une fois le GAM logistique binaire construit, les logits ajustés (« toutes choses étant égales par ailleurs », c'est-à-dire prenant en compte l'influence des autres variables) peuvent être visualisés pour chacune des variables explicatives du modèle (Figures 40 à 45). Ces graphes aident à l'interprétation de l'influence des variables sur la probabilité de présence de Pn. L'augmentation de la teneur en SiOH diminue globalement le risque d'apparition de Pn, tandis que celle de la richesse ou de la teneur en chlorophylle a l'augmente. La température et J augmentent ce risque jusqu'à une certaine valeur à partir de laquelle elles le diminuent. L'augmentation du débit de la Vilaine augmente le risque d'apparition de Pn.

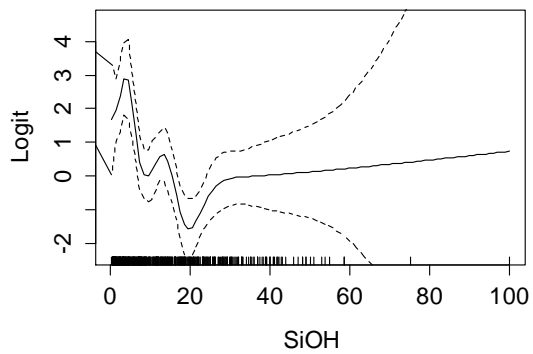


Figure 43 : Logit ajusté pour la teneur en SiOH

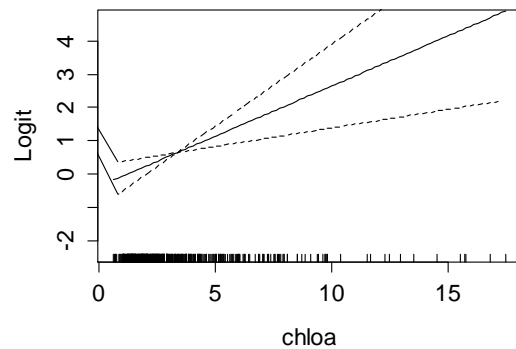


Figure 46 : Logit ajusté pour la teneur en chloa

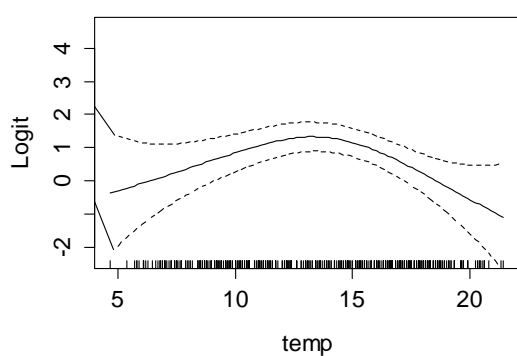


Figure 44 : Logit ajusté pour la température

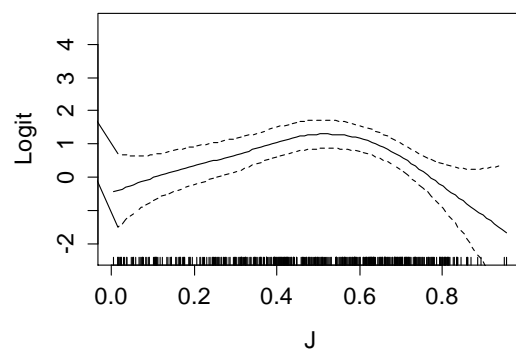


Figure 47 : Logit ajusté pour l'équitabilité de Piélu

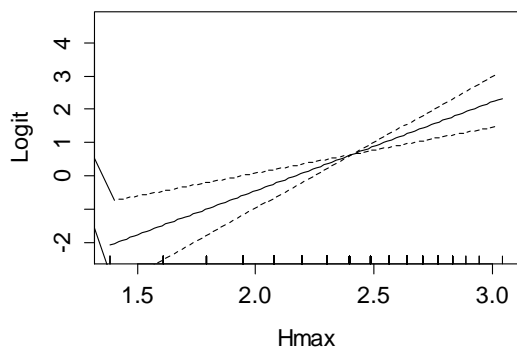


Figure 45 : Logit ajusté pour la richesse

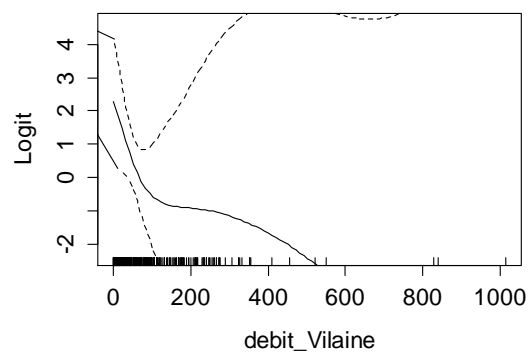


Figure 48 : Logit ajusté pour le débit de la Vilaine

IX.8.2. Annexe 8.2 : Le modèle expliquant la croissance de Pn

Les résidus du GAM expliquant la croissance de Pn sont normaux comme le confirme la Figure 49.

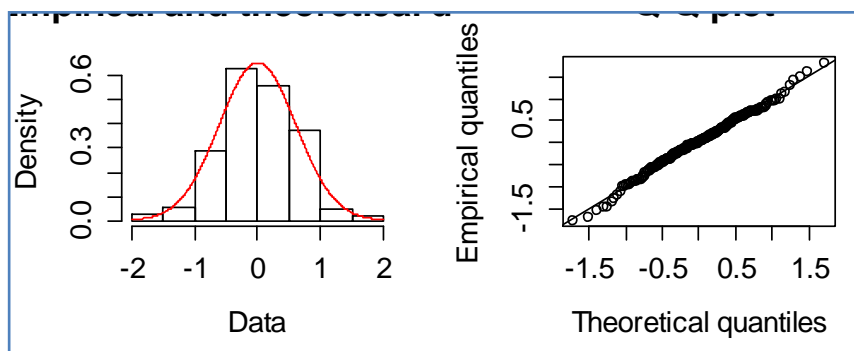


Figure 49 : Normalité des résidus du GAM expliquant la croissance de Pn

Il est également possible de visualiser l'effet de chaque variable du modèle en tenant compte des autres variables explicatives. Les résultats sont présentés aux Figures 49 à 53. Ces graphes aident à l'interprétation de l'influence de chacune des variables.

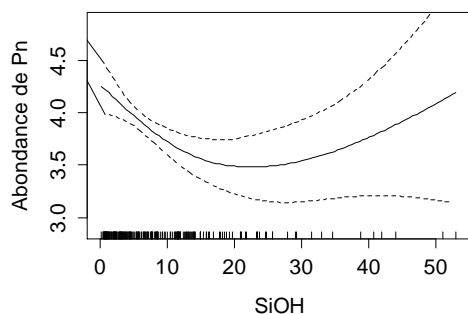


Figure 50 : Relation ajustée avec la teneur en SiOH

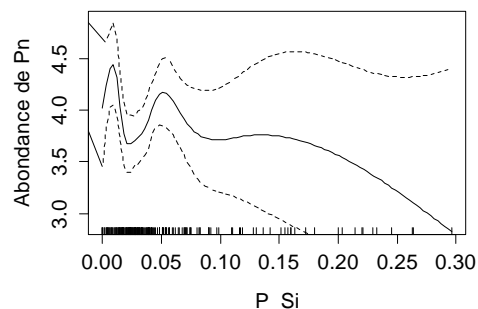


Figure 52 : Relation ajustée avec P/Si

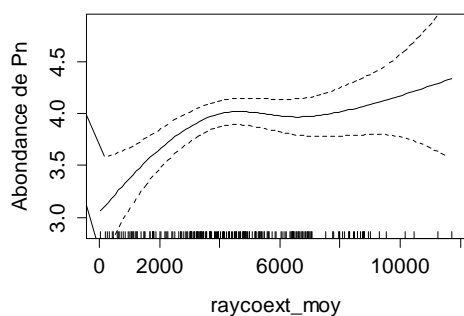


Figure 51 : Relation ajustée avec l'énergie utile moyenne (raycoext_moy)

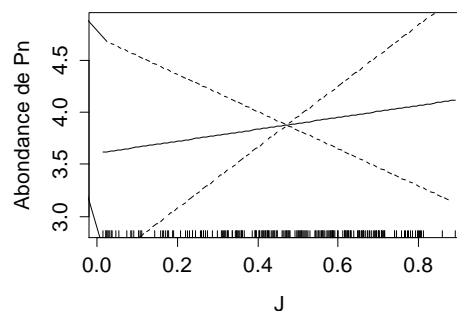


Figure 53 : Relation ajustée avec J

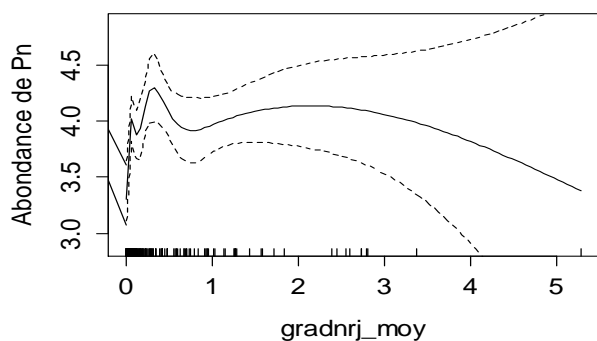


Figure 54 : Relation ajustée avec la stratification moyenne (gradnrj_moy)

La teneur en SiOH est corrélée négativement à l'abondance de Pn, en considérant la portion de courbe où l'intervalle de confiance est correct. L'énergie utile moyenne est au contraire corrélée positivement à l'abondance de Pn. Les autres relations restent difficiles à interpréter.

IX.9. Annexe 9 : Analyses complémentaires

IX.9.1. Annexe 9.1 : Une autre approche pour la modélisation multivariée

Une autre approche a été réalisée pour expliquer l'abondance de Pn en fonction des conditions environnementales. Cette approche est basée sur le caractère saisonnier de l'ensemble des variables de l'étude.

IX.9.1.1. Saisonnalité et tendance des variables environnementales

De même que pour l'abondance de Pn, la dépendance au temps des différentes variables environnementales a été étudiée. Deux variables sont présentées : la teneur en SiOH et la teneur en chlo a (ces deux variables sont par la suite sélectionnées dans les modèles explicatifs).

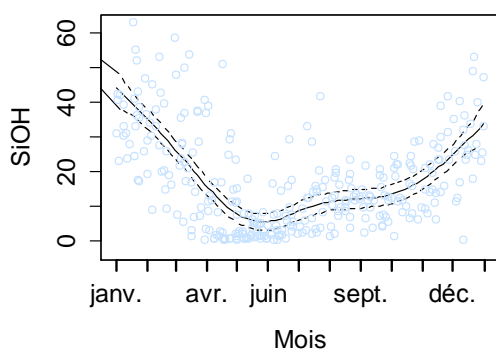


Figure 55 : Saisonnalité de la teneur en SiOH dans l'eau de mer à OL

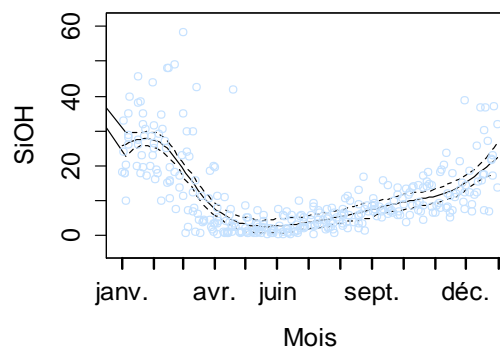


Figure 56 : Saisonnalité de la teneur en SiOH dans l'eau de mer à MeR

La saisonnalité est marquée et est plus ou moins semblable d'un site à l'autre : augmentation l'hiver puis diminution progressive jusqu'à une valeur minimale en été. On remarque que la teneur en silicates est globalement plus faible à MeR qu'à OL.

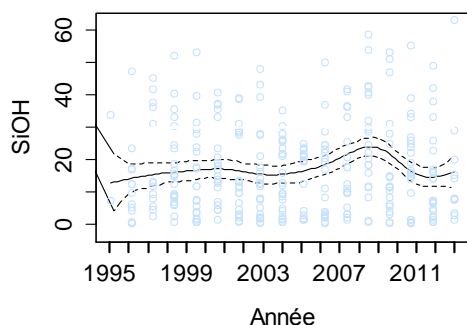


Figure 57 : Tendence de la teneur en SiOH dans l'eau de mer à OL

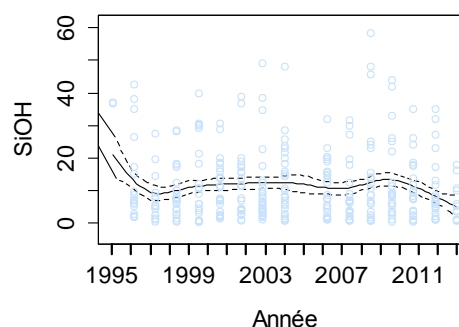


Figure 58 : Tendence de la teneur en SiOH dans l'eau de mer à MeR

La teneur en silicates dans l'eau de mer à OL ou MeR a peu évolué de 1995 à 2012. On remarque cependant une augmentation en 2009 à OL.

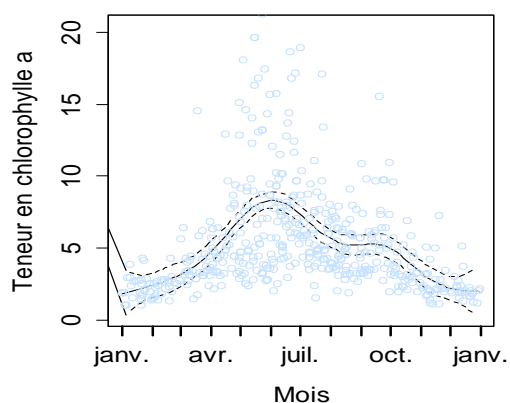


Figure 59 : Saisonnalité de la teneur en chlo a dans l'eau de mer à OL

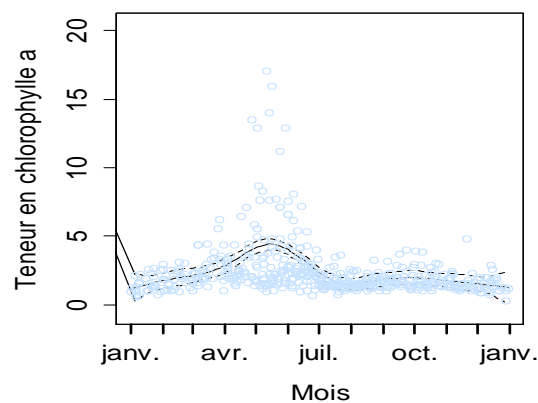


Figure 60 : Saisonnalité de la teneur en chlo a dans l'eau de mer à MeR

De même que pour les silicates, la **teneur en chlorophylle a** dans l'eau de mer est globalement **plus élevée à OL** qu'à MeR. Celle-ci augmente dès le printemps et diminue progressivement jusqu'à être très faible en hiver.

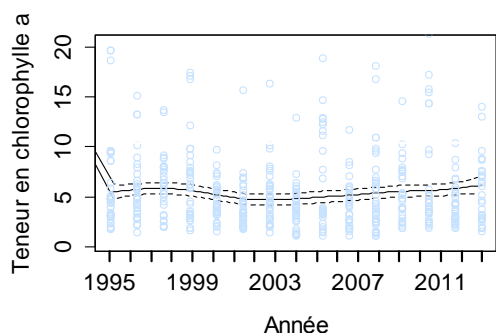


Figure 61 : Tendence de la teneur en chlo a de l'eau de mer à OL

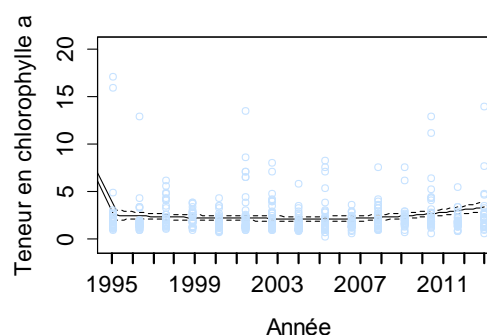


Figure 62 : Tendence de la teneur en chlo a de l'eau de mer à MeR

La teneur en chlorophylle a reste moyennement stable le long des années que ce soit à OL ou à MeR.

Des **variations temporelles** sont **constatées** de façon plus importante en cours d'année (saisonnalité) que le long des années (tendance), que ce soit à OL ou MeR et que ce soit pour l'abondance de Pn ou pour l'ensemble des variables environnementales (dont seuls deux exemples sont présentés).

IX.9.1.2. La décomposition de l'abondance de Pn

L'évolution temporelle de Pn est considérée comme une **moyenne saisonnière** : ce sont les valeurs d'abondance de Pn auxquelles on s'attend plus ou moins chaque année. La moyenne n'est donc pas constante (comme elle le serait dans une régression linéaire) mais dépendante du temps. La **variance** à expliquer correspond donc aux **fluctuations autour de cette moyenne**.

L'idée principale est donc de **décomposer** le « signal » d'abondance de Pn : une moyenne saisonnière et des fluctuations autour de cette moyenne. Ceci est effectué grâce au GAM en expliquant l'abondance de Pn en fonction du jour de l'année et de l'année pour chaque lieu (IV.2.2). La moyenne saisonnière est alors ce qui est prédit par un ce modèle, tandis que ses résidus sont les fluctuations autour de cette moyenne (Figure 63 et Figure 64).

Finalement, cette décomposition peut se résumer ainsi : **Valeurs réelles = Moyenne saisonnière + Résidus**. Chaque terme à droite de l'équation peut ensuite être **étudié séparément**.

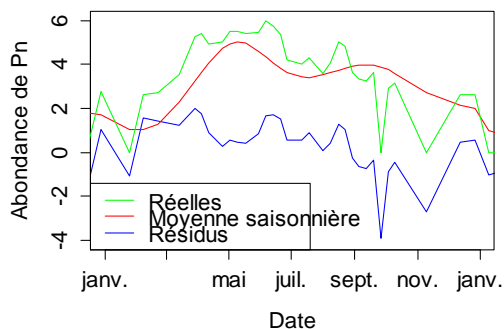


Figure 63 : Décomposition de l'abondance de Pn à OL en 2005

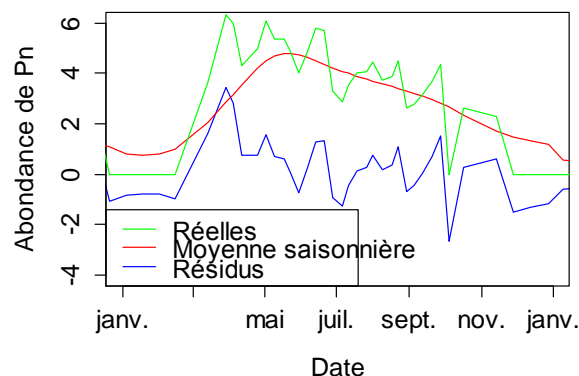


Figure 64 : Décomposition de l'abondance de Pn à MeR en 2005

Cette décomposition est effectuée également pour **chaque variable explicative**. La moyenne saisonnière de l'abondance de Pn est alors mise en relation avec les moyennes saisonnières des variables explicatives. De même, les fluctuations autour de la moyenne saisonnière de l'abondance de Pn sont expliquées par les fluctuations autour des moyennes saisonnières des variables explicatives.

Les modèles multivariés sont effectués de même que ceux dans le rapport : classification des variables explicatives selon leur influence en analyse bivariée (selon la significativité des coefficients estimés ou la variance expliquée par le modèle bivarié expliquant soit la moyenne saisonnière de l'abondance de Pn soit ses fluctuations autour de cette moyenne). Ensuite, chaque variable est introduite une à une (selon l'AIC ou le test du rapport de log vraisemblance).

IX.9.1.3. La moyenne saisonnière de l'abondance de Pn

On s'aperçoit qu'une fois « lissé », des relations plus « directes » avec l'abondance de Pn apparaissent (Figure 65 et Figure 66).

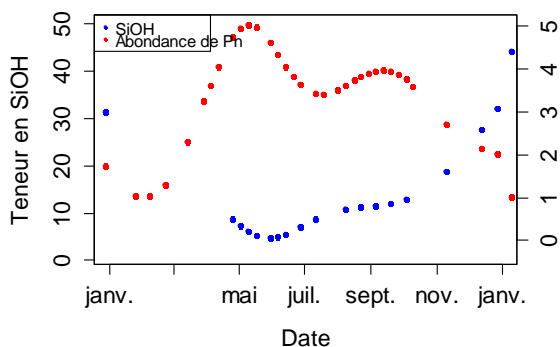


Figure 65 : Moyennes saisonnières de la teneur en SiOH et de l'abondance de Pn à OL en 2005

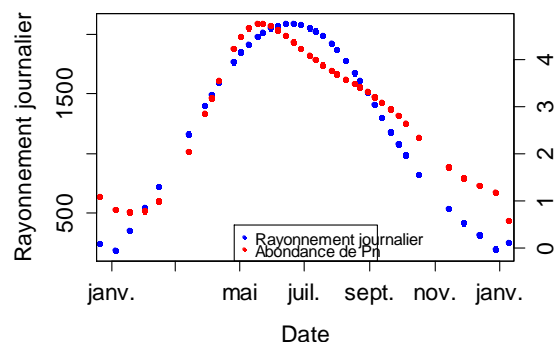


Figure 66 : Moyennes saisonnières du rayonnement journalier et de l'abondance de Pn à MeR en 2005

La moyenne saisonnière de l'abondance de Pn semble anti-corrélée tandis que celle du rayonnement journalier semble positivement corrélée à la moyenne saisonnière de l'abondance de Pn. Finalement, la moyenne saisonnière de l'abondance de Pn est modélisée par une régression linéaire. Deux variables explicatives sont retenues pour expliquer la moyenne saisonnière de l'abondance de Pn à OL : les moyennes saisonnières à OL des teneurs en silicates et en mies moyenne.

Tableau 11 : Paramètres de la régression linéaire expliquant la moyenne saisonnière de l'abondance de Pn à OL

	Coefficient estimé	Ecart-type	Statistique de test	Valeur de p	Indicateur de significativité
Ordonnée à l'origine	4,83	9,73E-02	49,628	2,00E-16	***
SiOH	-1,49E-01	8,44E-03	-17,674	2,00E-16	***
Teneur en mies moy.	-1,34E-02	1,27E-02	-1,057	2,92E-01	
SiOH : mies moy.	1,43E-03	3,13E-04	4,56	7,44E-06	***

Chaque paramètre est différent de zéro d'après les valeurs de p renseignés (tests de nullité des paramètres significatifs au risque 5%), excepté la teneur en smies moyenne, donc l'interaction avec la teneur en silicates est significative (Tableau 11). L'équation du modèle est donc :

$$\text{moy. sais. de l'abondance de Pn} = 4,83 - 0,149 * \text{moy.sais.SiOH} - 0,0134 * \text{moy.sais.Mies moy.} + 0,00143 * \text{moy.sais.SiOH : mies moy.}$$

D'après le R^2 ajusté, la variance expliquée par le modèle est de 84 %. Le modèle construit prédit mieux qu'un modèle nul (test significatif avec une valeur de p inférieure à $2,2 \cdot 10^{-16}$).

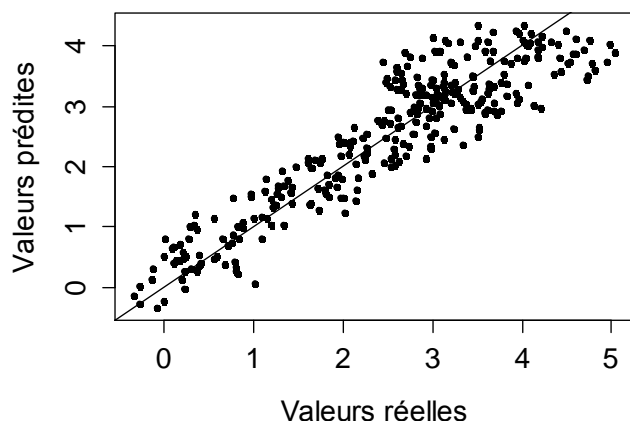


Figure 67 : Valeurs prédites par le modèle linéaire de la moyenne saisonnière de l'abondance de Pn à OL en fonction de ses valeurs réelles

Cette représentation suit plutôt bien la droite $y=x$: la prédiction du modèle est très bonne. De même, deux variables explicatives sont retenues pour expliquer la moyenne saisonnière de l'abondance de Pn à MeR : les moyennes saisonnières à MeR du rayonnement journalier et de la teneur en silicates.

Tableau 12 : Paramètres de la régression linéaire expliquant la moyenne saisonnière de l'abondance de Pn à MeR

	Coefficient estimé	Ecart-type	Statistique de test	Valeur de p	Indicateur de significativité
Ordonnée à l'origine	4,83	9,73E-02	49,628	2,00E-16	***
SiOH	-4,44E-02	5,32E-03	-8,342	2,01E-15	***
Rayonnement journalier	1,42E-03	6,25E-05	22,724	2,00E-16	***
SiOH :Rayon. jour.	-3,33E-05	4,84E-06	-6,88	3,01E-11	***

De même, chaque paramètre du modèle est significativement différent de zéro ($p < 0,05$). D'après le R^2 ajusté, la variance expliquée par le modèle est de 94 %. La prédiction du modèle est même meilleure que celui précédent (Figure 68).

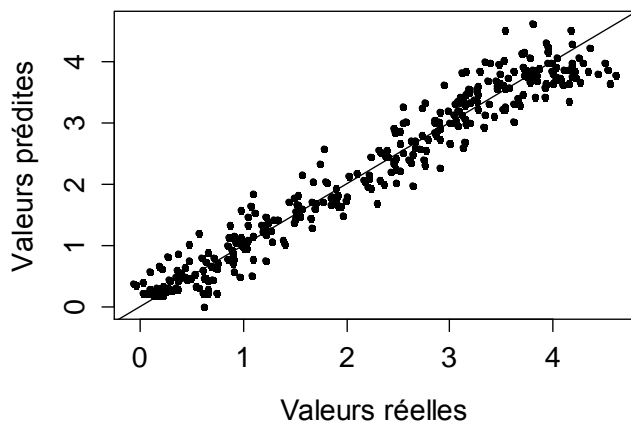


Figure 68 : Valeurs prédites par le modèle linéaire de la moyenne saisonnière de l'abondance de Pn à MeR en fonction de ses valeurs réelles

Les tests d'adéquation globale des modèles sont significatifs au risque 5 % (tests de Fisher avec $p=2.10^{-16}$). Les deux modèles sont de plus validés (Figure 69 et Figure 70) : les résidus sont normaux, leurs variances sont plus ou moins homogènes et la distance de Cook ne révèle pas de points très influents.

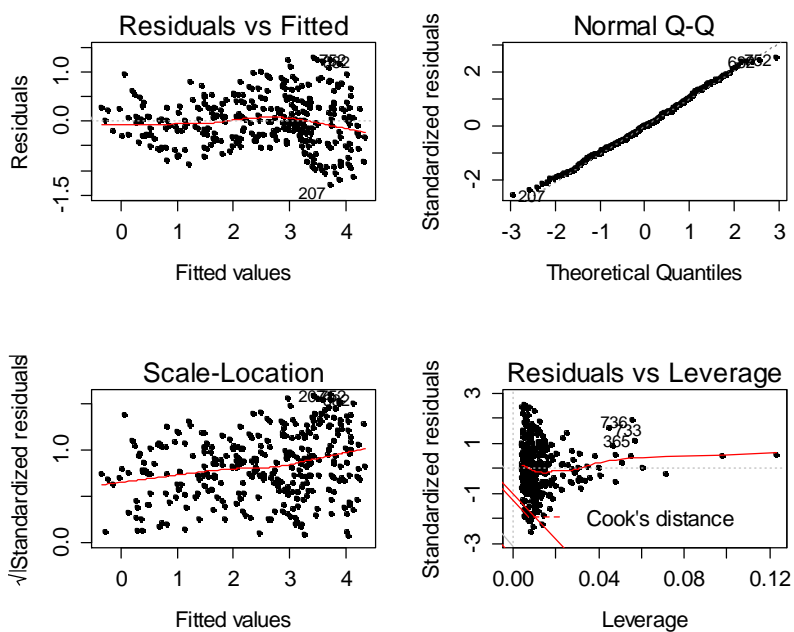


Figure 69 : Validation du modèle explicatif de la moyenne saisonnière de l'abondance de Pn à OL

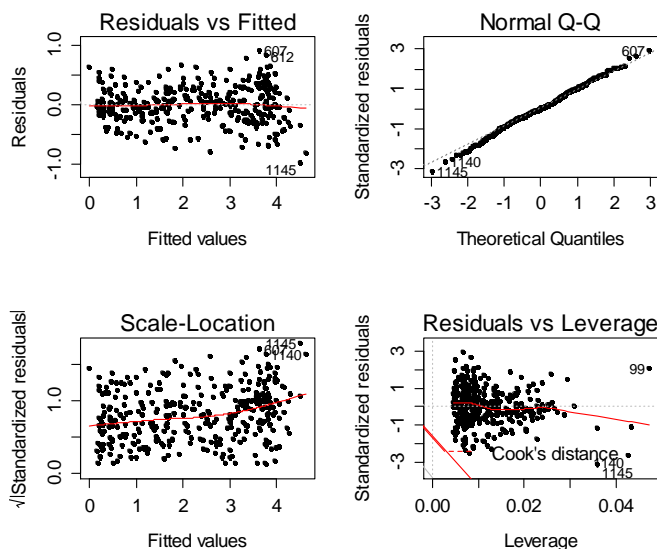


Figure 70 : Validation du modèle explicatif de la moyenne saisonnière de l'abondance de Pn à MeR

Les relations avec les moyennes saisonnières sont très bien mises en évidence. Celles avec les « résidus » sont plus complexes.

IX.9.1.4. Les fluctuations autour de la moyenne saisonnière de l'abondance de Pn

Cette fois, les relations ne sont pas linéaires : le GAM est utilisé pour expliquer la variance autour de la moyenne saisonnière. L'ordonnée à l'origine n'est pas significative : l'ordonnée à l'origine représente en fait une moyenne mais la moyenne a été extraite ici (c'est la moyenne saisonnière). Donc elle devait être égale à 0 dans ces modèles construits, ce qui est le cas (ordonnée à l'origine des modèles non significatives au risque 5%).

Quatre variables sont retenues pour expliquer les fluctuations de l'abondance de Pn autour de sa moyenne à OL : les teneurs en silicates et en phosphates, le rapport N/P et la stratification journalière.

Tableau 13 : GAM construit pour expliquer les fluctuations de l'abondance de Pn autour de sa moyenne saisonnière à OL

	Degré de liberté	Degré de liberté de référence	Statistique de test (F)	Valeur de p	Indicateur de significativité
SiOH	1,00	1,00	0,47	4,94E-01	
PO4	4,38	5,23	1,54	1,79E-01	
SiOH:PO4	6,46	27,00	0,98	3,68E-05	***
N/P	8,69	8,93	4,14	8,98E-05	***

Stratification	5,99	6,56	0,38	9,04E-01	
Stratif. :PO4	6,55	27,00	0,50	6,56E-03	**

La variance expliquée par le modèle est de 53%. La prédiction du modèle est moyenne (Figure 71).

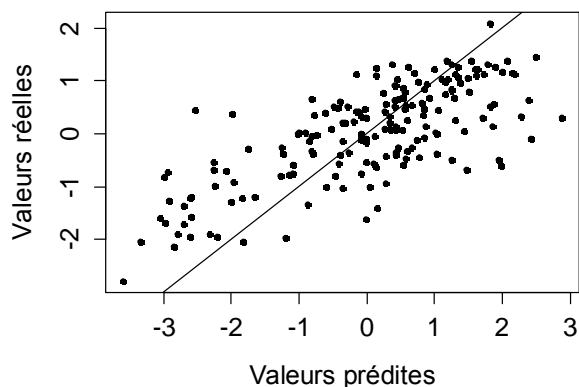


Figure 71 : Valeurs prédites des fluctuations autour de la moyenne saisonnière de l'abondance de Pn à OL en fonction des valeurs réelles

À MeR, cinq variables sont retenues pour expliquer les fluctuations autour de la moyenne saisonnière de l'abondance de Pn : les indices écologiques J et Hmax, le débit de la Loire moyen, la teneur en chlorophylle a et en phosphates (Tableau 14). Tous ces paramètres sont significatifs ($p < 0,05$).

Tableau 14 : GAM construit pour expliquer les fluctuations de l'abondance de Pn autour de sa moyenne saisonnière à MeR

	Degré de liberté	Degré de liberté de référence	Statistique de test (F)	Valeur de p	Indicateur de significativité
J	6,08	7,09	2,33	2,47E-02	*
Hmax	1,00	1,00	22,50	3,50E-06	***
Débit de la Loire moyen	1,44	1,75	4,07	2,39E-02	*
Teneur en chlo. A	1,25	1,45	3,93	3,40E-02	*
PO4	3,29	3,96	2,52	4,24E-02	*

La variance expliquée n'est que de 22,6 %. La prédiction du modèle est mauvaise (Figure 72).

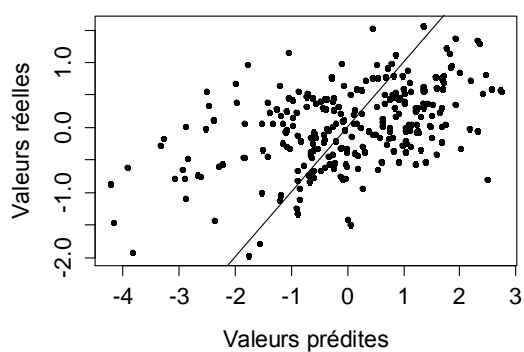


Figure 72 : Valeurs prédites des fluctuations autour de la moyenne saisonnière de l'abondance de Pn à MeR en fonction des valeurs réelles

Au niveau des résidus, les deux modèles précédents ne présentent pas de désaccords avec les hypothèses de normalité et de variance homogène (Figure 73 et Figure 74).

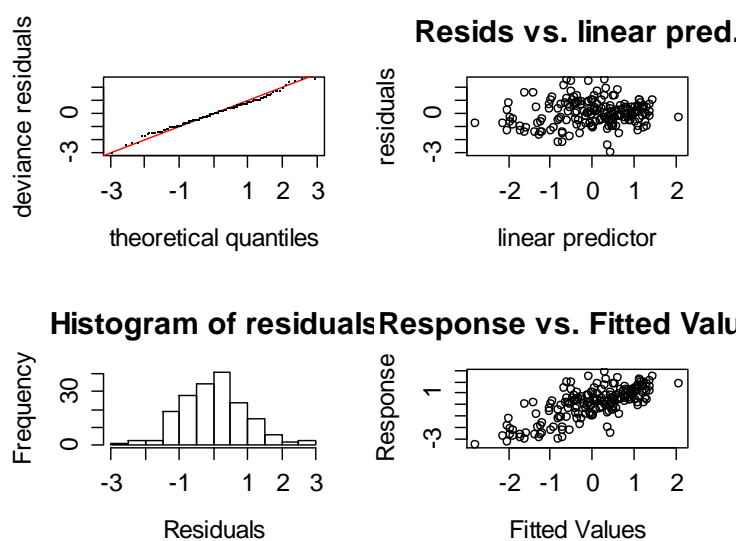


Figure 73 : Validation du GAM expliquant autour de la moyenne saisonnière de l'abondance de Pn à OL

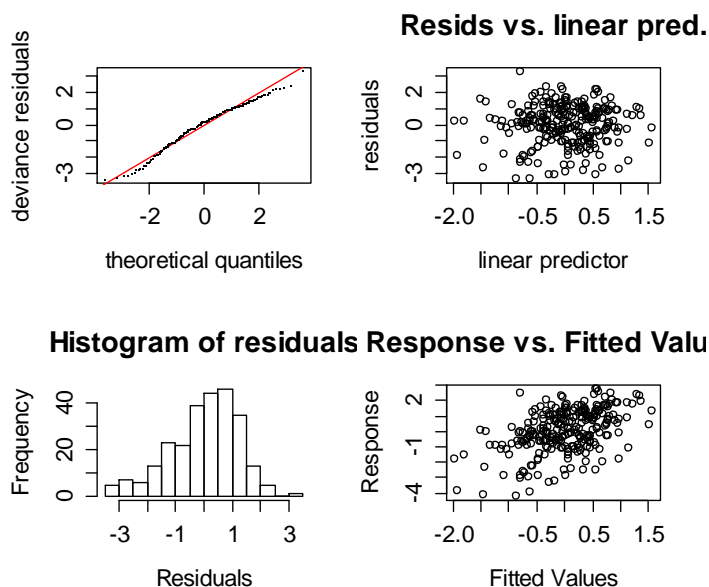


Figure 74 : Validation du GAM expliquant autour de la moyenne saisonnière de l'abondance de Pn à MeR

Des variables ont été sélectionnées pour expliquer les fluctuations autour de la moyenne saisonnière de Pn, mais le pouvoir prédictif des modèles est faible. Il manque sans doute des informations afin de mieux comprendre ces fluctuations et surtout celles positives (augmentant l'abondance attendu et le danger éventuel de toxicité).

Finalement, l'approche présentée a des avantages et des inconvénients. Elle a permis de travailler sur des données plus indépendantes : les modèles sont construits pour chaque lieu et la dépendance au temps est plus ou moins éliminée une fois la moyenne saisonnière extraite. Cependant, les modèles construits sont alors propres à chaque site (OL ou MeR) et les conclusions relatives à ces modèles ne peuvent pas être étendues à l'ensemble du Mor Braz.

La moyenne saisonnière contient la majeure partie de l'information sur l'abondance de Pn et elle peut être modélisée grâce à une régression linéaire (méthode plus interprétable que le GAM). Les résidus quant à eux, semblent néanmoins plus difficiles à expliquer, surtout à MeR.

La méthode présentée dans le rapport a l'avantage d'être plus simple à mettre en œuvre, mais toutes deux sont améliorables, et à améliorer.

IX.9.2. Annexe 9.2 : Analyse de la flore phytoplanktonique

Une ACP (Analyse en Composantes Principales) a été réalisée en prenant comme variables actives les abondances en log10 des ordres pertinents (rencontrés régulièrement et/ou toxiques) à

OL ou MeR ainsi que celle de Pn. L'ordre des Bacillariales comprenant le genre *P.nitzschia*, ce dernier en a été soustrait.

Elle montre globalement un effet taille (Figure 75) : plus un ordre est abondant, plus les autres le sont. Les variables sont toutes positivement corrélées à la dimension 1 (majorité de l'inertie, qui différencie les échantillons, représentée par cet axe).

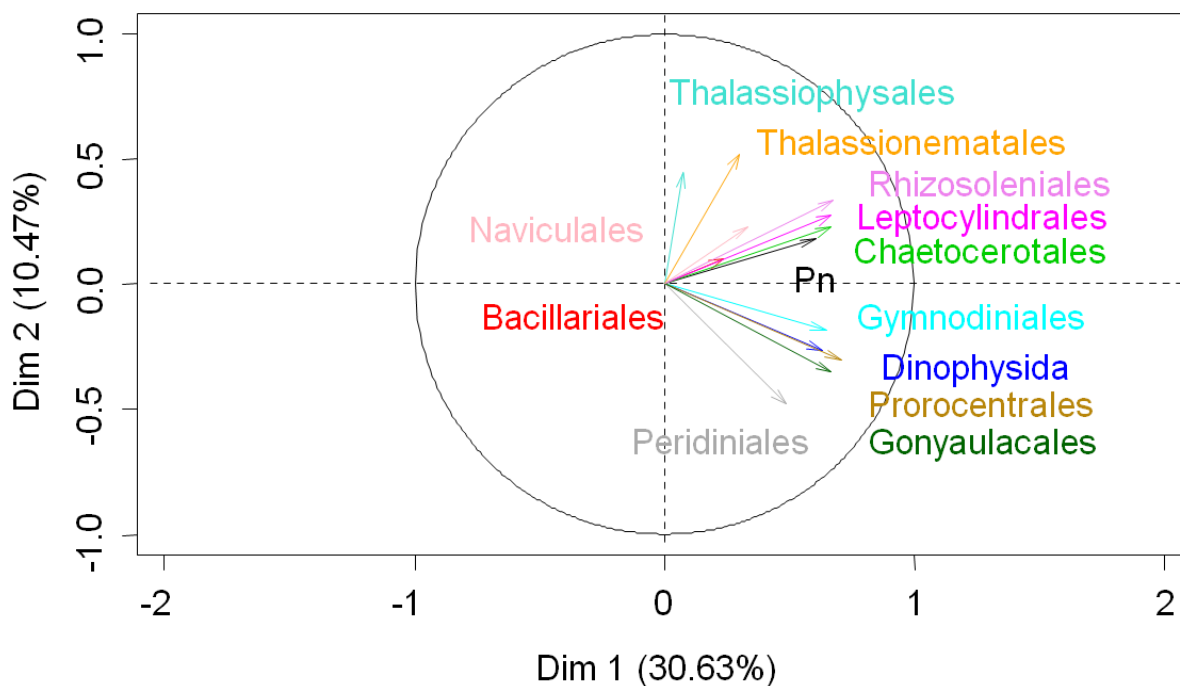


Figure 75 : Projection des variables actives sur les dimensions 1 et 2 de l'ACP

Une variable supplémentaire renseignant les saisons aide à l'interprétation. Ils sont plutôt abondants en été/printemps, opposés à l'hiver sur ces axes (Figure 77).

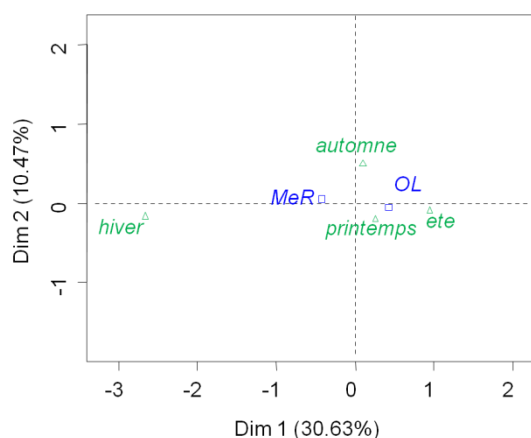


Figure 76 : Position des modalités des variables supplémentaires qualitatives sur les axes 1 et 2 de l'ACP

Les dimensions 3 et 4 représentent peu d'inertie mais séparent plus les ordres et saisons différentes de l'été (Figure 77 et Figure 78). Pn, les Rhizoleniales et Chaetocerotales sont plus abondants au

printemps et à MeR. Les Thalassiophysales ainsi que les Naviculales sont plutôt abondants en hiver et à OL. Les Bacillariales autres que Pn sont très opposées aux autres et semblent croître de préférence lorsque les autres sont peu abondants.

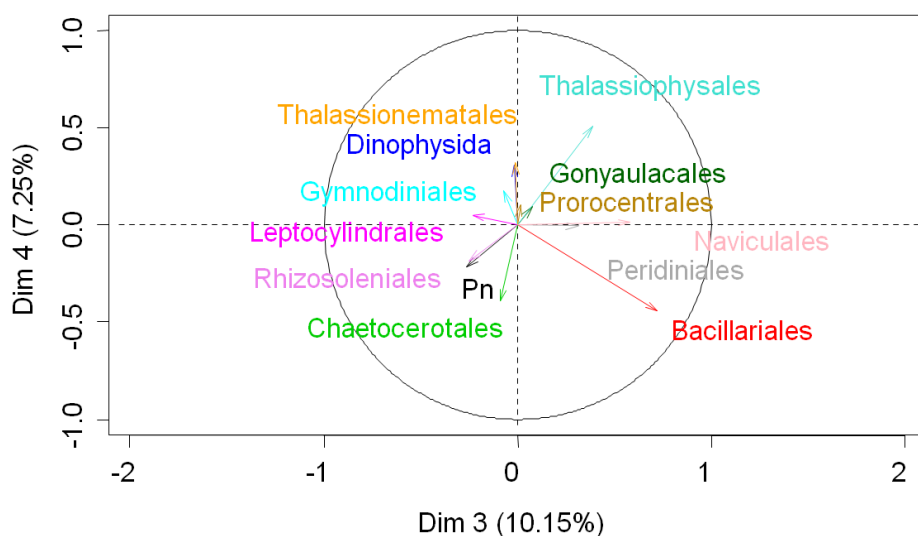


Figure 77 : Projection des variables actives sur les dimensions 3 et 4

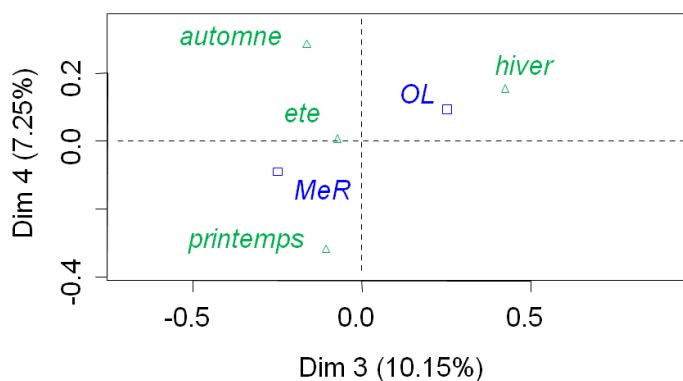


Figure 78 : Position des modalités des variables supplémentaires qualitatives sur les axes 3 et 4 de l'ACP

De la même manière, une ACP a été réalisée en prenant cette fois comme variables actives les proportions de chaque ordre, transformées en arcsin(\sqrt{v}). L'inertie n'est que faiblement représentée. La dimension 1 oppose les ordres « dominants » plutôt en hiver (à gauche de l'axe) des autres ordres (à droite) (Figure 79 et Figure 80).

On peut remarquer que la proportion/dominance de Pn est élevée lorsque celles des Rhizoleniales, Chaetocerotales et Leptocylindrales sont élevées (même rapprochement que pour l'ACP sur les abondances). Les Bacillariales autres que Pn sont encore une fois opposées à Pn.

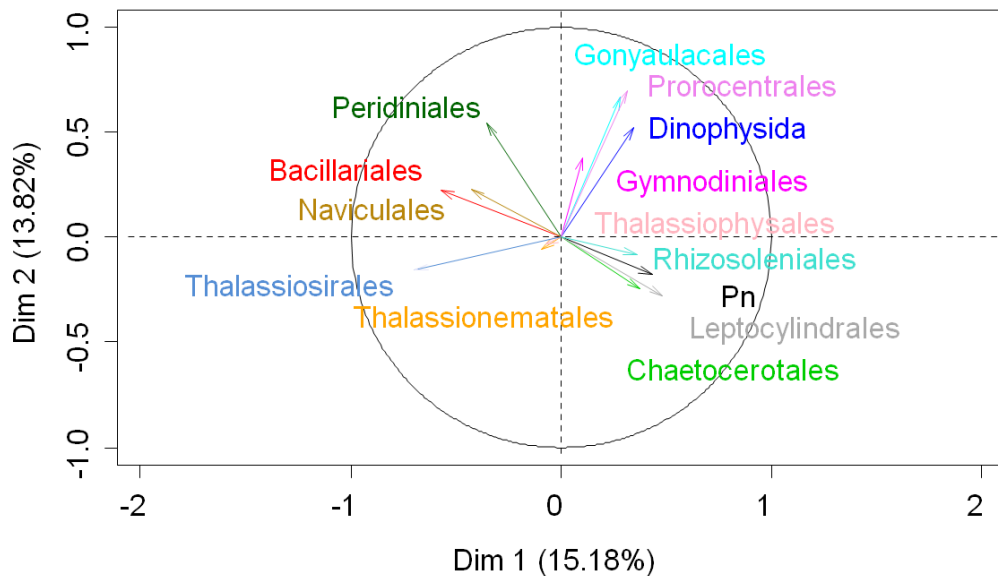


Figure 79 : Projection des variables actives sur les dimensions 1 et 2

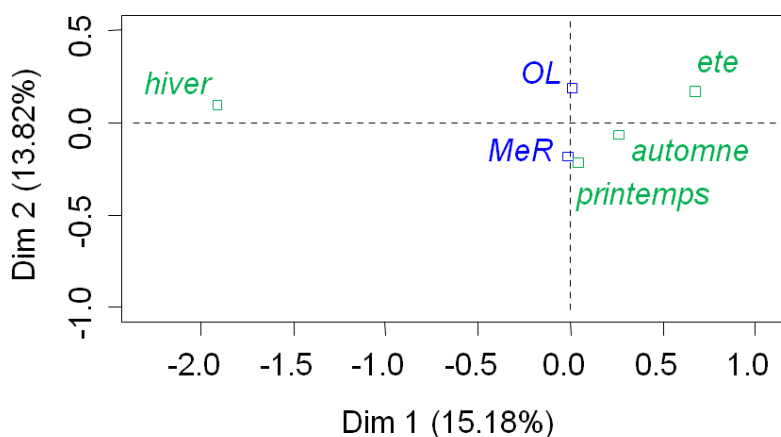


Figure 80 : Position des modalités des variables supplémentaires qualitatives sur les axes 1 et 2 de l'ACP

IX.9.3. Annexe 9.3 : Les conditions environnementales rencontrées à OL et MeR

Une ACP a été réalisée également pour décrire les conditions environnementales rencontrées dans les échantillons de la base à OL et MeR puis de vérifier si les disparités au niveau des conditions environnementales sont en relation avec les disparités d'abondance de Pn observées. Beaucoup de variables sont corrélées (Tableau 10). Ainsi, seules les variables à partir desquelles sont construites majoritairement les dimensions dans l'ACP et ayant des interprétations biologiques différentes sont conservées (ce qui évite la redondance d'informations et clarifie de plus la lecture des résultats). Les variables environnementales actives conservées dans l'ACP sont : la **chlorophylle a**, le **rayonnement moyen**, la **richesse**, l'**équitabilité de Piélu**, la teneur en **azote**, **silicates** et

phosphates et la teneur en **mies** moyenne de l'eau de mer. Les variables supplémentaires ajoutées à l'ACP sont l'**abondance de Pn**, le **lieu** (OL ou Mer), les **saisons** et les **dates de croissance maximale** (notées « pente max », « non » sinon, et obtenues grâce à la méthode d'analyse des efflorescences, III.4.2). L'abondance a également été mise en classe et ajoutée en variable supplémentaire dans l'ACP. Les classes sont réalisées comme indiquées dans le Tableau 15.

Tableau 15 : Abondance de Pn mise en classes

Abondance	0	10 000	100 000	1 000 000	-
Classes	0	1	2	3	4

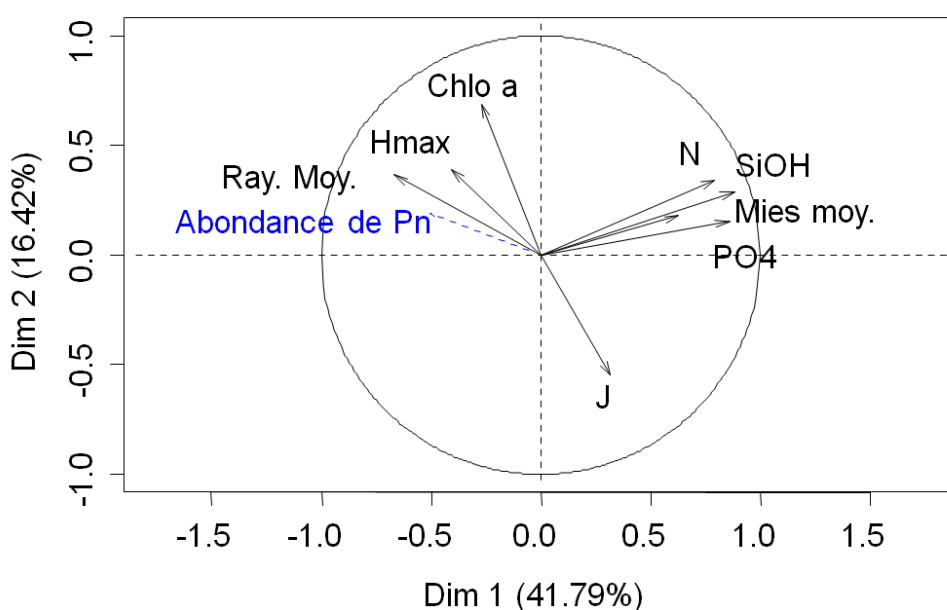


Figure 81 : Vecteurs des variables environnementales par rapport aux 2 premiers axes de l'ACP

La **dimension 1** représente les « **Conditions de croissance** » : elle résume principalement les informations relatives à la teneur en nutriments (PO4, N, et SiOH) et en mies moyenne de l'eau, et au rayonnement (dont les valeurs conditionnent la capacité de croissance ou non du phytoplancton). À droite de cette dimension (coordonnées positives sur la dimension 1), les conditions environnementales rencontrées peuvent être résumées ainsi : beaucoup de nutriments et de mies dans l'eau et peu de lumière (inversement à gauche de l'axe).

La **dimension 2** représente la « **Communauté phytoplanctonique** » : elle résume la diversité/richeesse (notée Hmax), l'équitabilité de Piélou (notée J) et la teneur en chlorophylle a. Elle décrit donc le phytoplancton dans son ensemble pour chaque échantillon prélevé. À droite de cet axe (coordonnées positives sur la dimension 2), les échantillons comportent beaucoup de chlorophylle a,

une richesse plutôt élevée mais une équitabilité de Pielou plutôt faible (inversement à gauche de cet axe).

L'**abondance de Pn** est bien corrélée à la **dimension 1** : l'abondance de Pn semble donc bien conditionnée par les nutriments, la turbidité de l'eau (la teneur en spim) et le rayonnement. La communauté phytoplanctonique semble n'avoir qu'une faible influence.

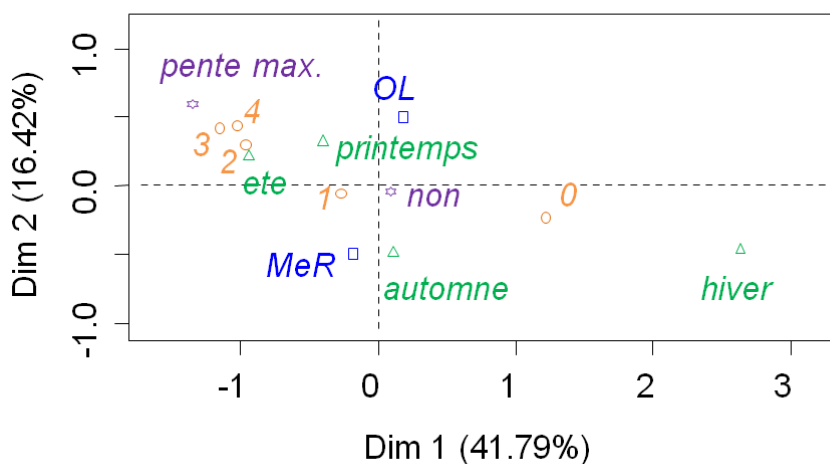
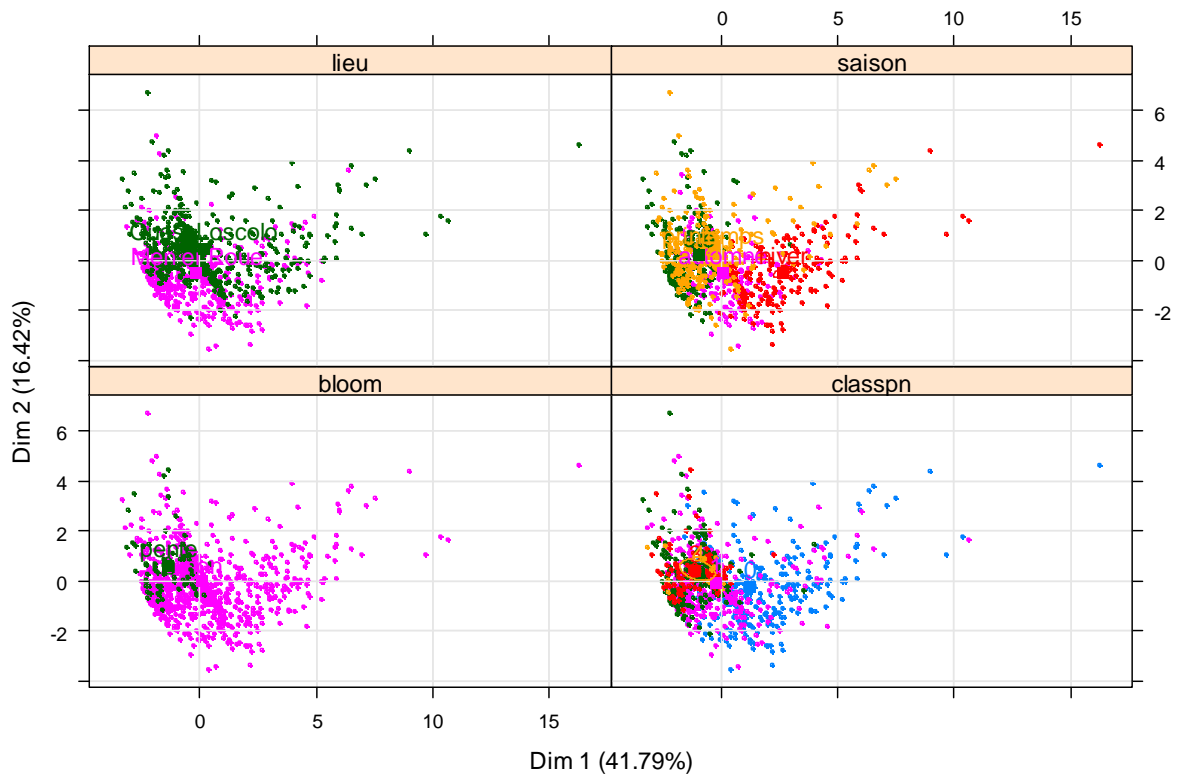


Figure 82 : Positionnement des modalités des variables qualitatives par rapport aux dimensions 1 et 2 de l'ACP

Les dimensions 1 et 2 décrivent relativement bien les **saisons** : elles sont plutôt bien réparties les une par rapport aux autres sur ces axes. L'hiver est notamment très opposé au printemps et à l'été. Les échantillons prélevés aux dates de croissance optimale de Pn ont une position éloignée des autres échantillons par rapport à ces axes : ils décrivent donc plutôt bien ces échantillons par rapport aux autres. La dimension 2 sépare moyennement les échantillons prélevés à OL de ceux prélevés à MeR.



La représentation via la mise en classe des valeurs de l'abondance de Pn indique que seules les échantillons avec une **abondance de Pn nulle** sont bien **séparés** des autres échantillons par rapport aux axes de l'ACP. En effet, les échantillons avec une abondance de Pn nulle ont en général des coordonnées positives par rapport à la dimension 1 (avec dans l'ensemble beaucoup de nutriments et peu de lumière dans l'eau) tandis que les échantillons avec une abondance de Pn supérieure à 10 000 cellules/L ont généralement des coordonnées négatives par rapport à cette même dimension (avec dans l'ensemble peu de nutriments mais beaucoup de lumière pénétrant dans l'eau)ⁱ.

La classification ne fournit que peu d'informations par rapport à l'abondance de Pn. Elle permet de mettre en valeurs des **profils de conditions environnementales** différentes rencontrées à OL ou MeR. Cette analyse est mise en annexe.

Des analyses supplémentaires sont nécessaires pour mieux décrire l'abondance de Pn par rapport aux variables environnementales : on utilise ici la modélisation.

ⁱ Construire l'ACP uniquement sur les données en cas de présence de Pn ne révèle pas plus d'informations.

Tableau 16 : Corrélations entre les dimensions de l'ACP et les variables quantitatives

	Dimension 1		Dimension 2	
	Corrélation	Valeur de p	Corrélation	Valeur de p
SiOH	0,883	0,00	0,284	0,00
PO4	0,862	0,00	0,154	1,13E-07
N	0,796	0,00	0,340	0,00
Spim moyen	0,629	0,00	0,176	1,21E-09
J	0,317	0,00	-0,552	2,25E-94
Chloro. A	-0,270	4,83E-21	0,688	0,00
Hmax	-0,410	1,17E-48	0,390	0,00
Ray. Moy.	-0,672	1,22E-154	0,364	0,00
Abondance de Pn	-0,508	7,96E-78	0,190	6,09E-11

Les corrélations les plus fortes avec la dimension 1 (en bleu) sont celles de la teneur en nutriments (PO₄, SiOH et N) et de la teneur en spim de l'eau (corrélations positives) ainsi que du rayonnement moyen (corrélations négatives). La dimension 1 est moyennement corrélée à l'abondance de Pn. Pour la dimension 2 de l'ACP, les 3 corrélations les plus fortes (en vert) sont : la teneur en chlorophylle a (corrélations positive) et l'équitabilité de Piéluou notée J (corrélations négative) ainsi que la richesse notée Hmax (qui est représentée tout autant par la dimension 1).

Seule la saison est moyennement expliquée par l'ACP (à environ 44% d'après le R² obtenu), surtout par la dimension 1 (Tableau 17).

Tableau 17 : Variable qualitatives expliquées par l'ACP

	Dimension 1		Dimension 2	
	R2	Valeur de p	R2	Valeur de p
Saison	0,436	9,36E-145	0,098	6,86E-26
Bloom	0,036	5,75E-11	0,018	4,09E-06
Lieu	0,010	5,68E-04	0,187	1,49E-54

Le Tableau 18 donne les coordonnées moyennes estimées des modalités des variables qualitatives supplémentaires incluses dans l'ACP. Pour chacune de ces estimations un test de nullité

de cette estimation est réalisé : les valeurs de p indiquées correspondent aux résultats de ces tests. Toutes ces coordonnées sont donc significativement différentes de zéro au risque 5% ($p < 0,05$).

Tableau 18 : Positionnement des catégories des variables qualitatives par rapport aux axes de l'ACP

	Dimension 1		Dimension 2	
	Coordonnée estimée	Valeur de p	Coordonnée estimée	Valeur de p
Hiver	2,284	1,56E-129	-0,362	4,18E-08
Non	0,718	5,75E-11	-0,318	4,09E-06
OL	0,184	5,68E-04	0,495	1,49E-54
Automne	-0,242	1,21E-03	-0,380	2,05E-10
MeR	-0,184	5,68E-04	-0,495	1,49E-54
Pente	-0,718	5,75E-11	0,318	4,09E-06
Printemps	-0,752	1,92E-25	0,422	8,67E-14
Été	-1,290	3,34E-78	0,320	3,54E-10

Ce sont des compléments par rapport à ce qui est vu graphiquement. Exceptée l'automne, les saisons sont assez bien séparées par la dimension 1, qui oppose aussi moyennement bien les échantillons prélevés aux dates de conditions de bloom de ceux aux autres dates. La dimension 2 sépare moyennement les échantillons prélevés à OL de ceux prélevés à MeR.

En observant la variable abondance de Pn au sein des axes de l'ACP, des informations sur l'abondance de Pn en fonction des conditions environnementales ont pu être acquises. Sur les axes de l'ACP, les conditions générales d'abondance de Pn nulle sont opposées à celles d'abondance supérieure à 10 000 cellules/L. L'ACP permet donc de décrire les **conditions environnementales générales de croissance de Pn et d'apparition des efflorescences : peu de nutriments** (SiOH, PO4 et N), **beaucoup de lumière** (rayonnement élevé) et une **eau claire** (avec une teneur en particules en suspension faible), ainsi qu'une équitabilité de Piélou plutôt faible, une diversité du phytoplancton et une teneur en chlorophylle a plutôt élevées. Ces conditions correspondent à celles rencontrées en général au printemps et en été et sont opposées à celles rencontrées en hiver.

Une classification hiérarchique ascendante avec consolidation des classes par la méthode des moyennes mobiles a été réalisée suite à l'ACP. Elle se base sur les huit premières dimensions de l'ACP (inertie cumulée de 100 %). Les classes obtenues sont représentées sur la Figure 83.

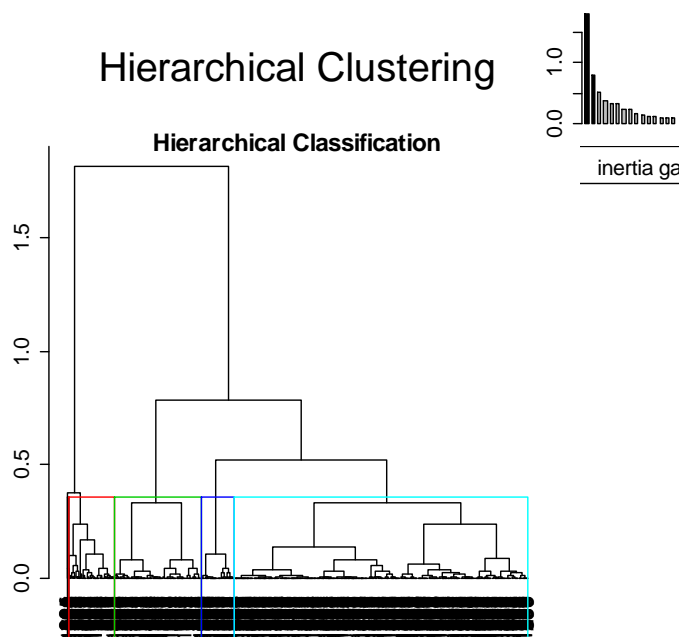


Figure 83 : Inertie donnée par les dimensions de l'ACP et dendrogramme de la classification effectuée

5 classes sont constituées. La classe 1 comporte 93 échantillons, 699 échantillons pour la classe 2, 272 pour la classe 3, 95 pour la classe 4 et enfin 13 pour la classe 5. Chaque classe est représentée selon les axes 1 et 2 de l'ACP sur la Figure 84.

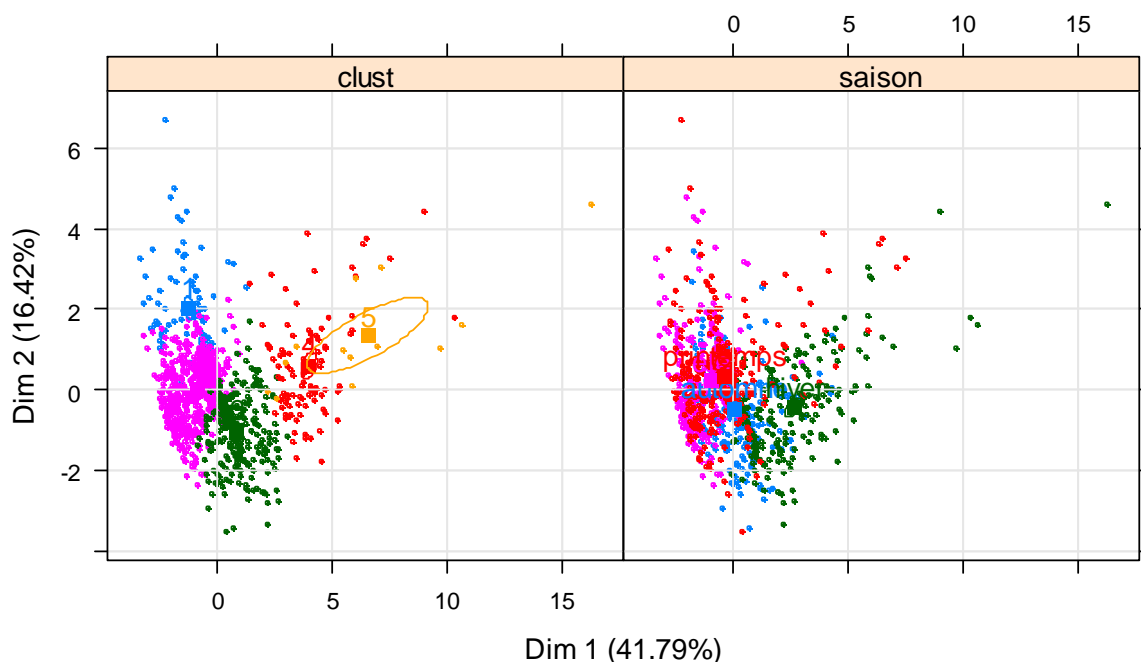


Figure 84 : Représentations des classes sur le plan des dimensions 1 et 2 de l'ACP

Le Tableau 19 résume les caractéristiques de chaque classe.

Tableau 19 : Classes obtenues ordonnées par rapport aux moyennes de chaque variable pour chaque classe (des plus forte aux plus faibles moyennes)

	+				-
Rayonnement moyen	1	2	3	4	5
Chlorophylle a	1	2	4	3	5
Abondance de Pn	1	2	3	5	4
Richesse	2	4	1	3	5
Phosphates	5	4	3	1	2
Equitabilité de Piélou	5	3	4	2	1
Teneur en spim	5	4	3	2	1
Teneur en SiOH	5	4	3	1	2
Teneur en N	4	5	3	1	2

La classe 1 est caractérisée par de fortes valeurs de rayonnement, de chlorophylle a et d'abondance de Pn. La diversité du phytoplancton est moyenne dans la classe 1 par rapport aux autres classes. Elle a de faibles valeurs d'équitabilité de Piélou et de teneur en spim et en nutriments.

La classe 2 est plus ou moins semblable à la classe 1, mais dans de moindres mesures (le rayonnement est par exemple beaucoup plus élevé dans la classe 2 en moyenne). Cependant l'abondance de Pn est à peu près semblable en moyenne que celle dans la classe 1. La richesse dans cette classe est a plus élevé en moyenne.

La classe 3 est une classe moyenne par rapport aux autres.

La classe 4 a de faibles valeurs de rayonnement, chlorophylle a et d'abondance de Pn. Ses valeurs de richesse, teneur en nutriments et en spim sont plutôt élevées.

La classe 5 est l'opposé de la classe 1 : elle détient des valeurs extrêmes de nutriments et de spim et de faibles valeurs de rayonnement, chlorophylle a, richesse et abondance de Pn.

91,4% des échantillons de la classe 1 sont prélevés en printemps/été, ce qui ne correspond néanmoins qu'à 22% des données de printemps/été de la base exploitée. Les autres sont prélevés en automne. 87% des échantillons de la classe 1 sont prélevés à OL. Cette classe contient 20% des échantillons prélevés lors des fortes croissances de Pn identifiées dans la partie

La classe 2 contient le plus d'effectifs. Elle contient 87% des prélèvements d'été de la base et 71% des prélèvements de printemps ainsi que 40% de ceux d'automne. Elle a environ autant d'échantillons de MeR que de OL. 76,7% des échantillons prélevés lors des fortes croissances de Pn sont dans cette classe.

49% des échantillons de la classe 3 ont été prélevés en automne et 37,5% en hiver. 60% des échantillons de cette classe ont été prélevés à MeR. Elle contient les derniers échantillons prélevés au moment des fortes croissances de Pn identifiées.

La classe 4 contient 70% de données d'hiver et 7% d'automne.

La classe 5 ne contient que peu d'effectifs : 12 échantillons de la classe 5 (qui en contient 13) ont été prélevés à OL. 11 ont été prélevés en hiver, 1 au printemps et 1 en été. (conclure globalement)

En ce qui concerne l'abondance de Pn, les classes ne se différencient pas toutes distinctement. 2 catégories peuvent être retenues au final : une première avec les classes 1 et 2 où l'abondance de Pn est globalement supérieure à 1000 cellules et une autre catégorie avec les classes 3, 4 et 5 où les abondances de Pn sont généralement inférieures à 1000 cellules. Ces 2 catégories s'opposent par rapport à des conditions de croissance de Pn (plutôt favorable dans la première catégorie que dans le 2^{ème}).