

INCERTITUDES DES MÉTHODES D'ÉVALUATION « EAUX LITTORALES » : UTILISATION DE MODÈLES LINÉAIRES DYNAMIQUES POUR L'ÉVALUATION DES INCERTITUDES (CHLOROPHYLLE A, PHYTOPLANCTON)

**Transférer les méthodes de bio indication.
Compléments apportés au guide méthodologique et
EIL phytoplancton et participation au CST**

**Soudant D., Miossec L., Neaud-Masson N., Auby I.,
Maurer D., Daniel-Scuiller A.**

Programme scientifique et technique
Année 2015

Document final



Contexte de programmation et de réalisation

Ce rapport a été réalisé dans le cadre du programme d'activité AQUAREF pour l'année 2015.

Auteur (s) :

Soudant Dominique
Ifremer centre de Nantes
dsoudant@ifremer.fr

Laurence Miossec
Ifremer centre de Nantes
lmiossec@ifremer.fr

Nadine Neaud-Masson
Ifremer centre de Nantes
Nadine.Masson@ifremer.fr

Isabelle Auby
Ifremer, laboratoire environnement ressource Arcachon
isabelle.auby@ifremer.fr

Danièle Maurer
Ifremer, laboratoire environnement ressource Arcachon
daniele.maurer@ifremer.fr

Anne Daniel-Scuiller
Ifremer, laboratoire DYNECO/PELAGOS
anne.daniel@ifremer.fr

Les correspondants

Onema : Yorick Reyjol

Etablissement : Dominique Soudant

Référence du document : Dominique Soudant, Laurence Miossec, Nadine Neaud-Masson, Isabelle Auby, Danièle Maurer, Anne Daniel-Scuiller - Incertitudes des méthodes d'évaluation « eaux littorales » : utilisation de modèles linéaires dynamiques pour l'évaluation des incertitudes (chlorophylle a, phytoplancton) - Rapport AQUAREF 2015 - 42 p.

Droits d'usage :	<i>Accès libre</i>
Couverture géographique :	<i>International</i>
Niveau géographique :	<i>National</i>
Niveau de lecture :	<i>Professionnels, experts</i>
Nature de la ressource :	<i>Document</i>

1. INTRODUCTION.....	6
2. DONNÉES ET MÉTHODES.....	7
2.1 Réseau, lieux, période, méthodes de mesures.....	7
2.2 Informations exogènes, données extraites et préparation.....	8
2.3 Méthodes.....	10
3. RÉSULTATS ET EXPERTISES.....	17
3.1 Chlorophylle <i>a</i>	17
3.2 Abondance phytoplanctonique.....	17
4. DISCUSSION.....	21
5. CONCLUSION.....	26
6. BIBLIOGRAPHIE.....	26
7. ANNEXE 1 : TABLE DE CORRESPONDANCE TAXONS-PHYLUM.....	29
8. ANNEXE 2 : DÉTAILS DES EXPERTISES.....	32
8.1 Chlorophylle <i>a</i>	32
8.2 Abondance phytoplanctonique.....	37

RESUMÉ

L'incertitude associée à une mesure a pour origine d'une part la variabilité environnementale et d'autre part l'ensemble du processus d'acquisition depuis le prélèvement jusqu'à la saisie de la donnée dans une base. L'estimation de l'ensemble de cette variabilité est un exercice complexe à réaliser dans le cadre d'un plan d'expérience. En revanche, les séries temporelles présentent la caractéristique d'intégrer toutes les variabilités et ainsi l'analyse de ces séries en terme de signal et bruit doit permettre de quantifier l'amplitude des incertitudes. Toutefois, les séries temporelles d'observation présentent un ensemble de caractéristiques les rendant difficiles à analyser. Les modèles linaires dynamiques constituent une approche adaptée à ces données particulières en faisant l'hypothèse de paramètres variables dans le temps. Ainsi, l'objet du présent travail consiste à estimer les variances liées au processus d'observation à l'aide de modèles linéaires dynamiques. Plus particulièrement, les mesures considérées sont la chlorophylle *a* et l'abondance phytoplanctonique aux lieux de surveillance REPHY « Arcachon-Bouée- 7 » et « Teychan bis ».

Les résultats montrent que pour la chlorophylle *a*, la variabilité d'observation est responsable de l'ordre de 80 % de la variabilité totale. Pour l'abondance phytoplanctonique, elle est également de 80 % à « Arcachon-Bouée 7 » mais de l'ordre de 70 % à « Teychan bis ». Ainsi la part de « bruit » est liée au lieu et au paramètre considéré. Exprimée en pourcentage de la médiane de la chlorophylle *a*, la variance d'observation place les bornes de l'intervalle de confiance à 95 % des observations à des valeurs de l'ordre de -40 % et +120 % de la médiane, l'intervalle étant sous estimé car ne prenant pas en compte la variabilité structurelle. Pour l'abondance phytoplanctonique en \log_{10} cell./L, les ordres de grandeur correspondant en pourcentage de la moyenne sont de ± 13.5 %. Pour les deux paramètres, ces valeurs sont compatibles avec l'expérience des experts.

Ainsi, l'approche mise en œuvre s'est avérée riche d'enseignements en matière d'incertitude de mesure et les nombreuses améliorations méthodologiques envisagées ouvrent des perspectives fécondes à tout point de vue.

Mots clés (thématique et géographique) :

Incertaince, chlorophylle *a*, abondance phytoplanctonique, modèle linéaire dynamique, Arcachon, REPHY.

1. INTRODUCTION

Le processus d'acquisition d'une mesure commence par le prélèvement et se termine par la saisie dans une base de données. Tout au long de ce processus, des phénomènes naturels et des opérations techniques sont susceptibles d'affecter le résultat final. Ainsi, le même processus réalisé plusieurs fois pour l'obtention de la même mesure ne conduit généralement pas au même résultat. Par voie de conséquence, il y a lieu de considérer l'existence d'une variabilité attachée à l'ensemble du processus d'acquisition d'une donnée. Par convention, appelons cette variabilité la variance d'observation.

Le processus d'acquisition peut être scindé en deux parties : une première concernant les opérations de prélèvement et une seconde les opérations de laboratoire. Dans la première partie, la variabilité environnementale, va contribuer de manière importante à la variance d'observation. La variabilité environnementale recouvre ici tout à la fois la variabilité spatiale et temporelle pertinente à l'échelle du prélèvement en ce qui concerne la quantité étudiée mais également l'ensemble des variations spatiales et temporelles de l'environnement de cette quantité et susceptible de l'affecter. Les distributions spatio-temporelles de la quantité étudiée et de son environnement sont corrélées mais pas nécessairement identiques. Par ailleurs, tout le protocole de prélèvement comporte nombre d'opérations (e.g. préparation, réalisation, conditionnement, conservation, transport) à même d'engendrer une variabilité dans le résultat. Dans la seconde partie, une fois au laboratoire, l'échantillon est également l'objet de multiples manipulations (e.g. sous échantillonnage, filtration, adjonction de réactifs et/ou de fixateur, quantification via du matériel d'analyse spécialisé ou comptage au microscope, conversions et calculs supplémentaires) qui sont potentiellement autant de nouvelles sources de variabilité jusqu'à et y compris la saisie informatique. Cette dernière étape ne marque pas la fin de la vie de la mesure, et ainsi la fin des sources de variabilité, puisque des processus de qualification, de révision, d'analyse et valorisation des données acquises sont susceptibles d'entraîner des changements dans les valeurs bancarisées. Toutefois, l'objet d'intérêt étant ici l'incertitude des mesures, ces ultimes étapes dans le processus d'observation ne seront pas considérées.

De nombreuses études ont été menées en ce qui concerne la variabilité environnementale. Les dimensions spatiales et temporelles sont de première importance. Les phénomènes responsables sont variés (Venrick, 1978a) et induisent de multiples échelles pertinentes (Dickey, 2003). Ces études ont parfois mené à une quantification de la variabilité environnementale (Cloern et al., 1989; Gagnon and Lacroix, 1981). Ces approches sont très liées à des questionnements concernant le processus de prélèvement à proprement parler sans toutefois en aborder toutes les dimensions. De plus, l'étude de la variabilité environnementale et celle induite par les opérations en laboratoire sont souvent déconnectées. Ainsi des études se concentrent sur l'impact du sous échantillonnage (Venrick, 1978b), des réactifs utilisés (Zarauz and Irigoien, 2008) ou des procédures de comptage (Venrick, 1978c).

L'analyse de variance est un outil statistique particulièrement bien adapté à l'étude des sources de variabilité et leur quantification (Alverson et al., 2003). Toutefois, son application se fait dans le cadre de plans d'expériences visant à contrôler certaines sources de variabilité afin de mieux étudier celles faisant l'objet de l'étude. Elle repose sur des hypothèses qui sont rarement vérifiées pour les données environnementales marines mais, d'une part, elle est très robuste aux écarts à ces hypothèses et, d'autre part, des transformations de variable permettent de satisfaire aux conditions d'application. Pour des raisons opérationnelles, il est difficile d'envisager un plan d'expérience ayant pour ambition d'embrasser la totalité du processus d'acquisition d'une mesure, de son

prélèvement à sa saisie. Les essais inter-laboratoires constituent un cadre dans lequel l'ensemble de la variabilité est parfois intégrée mais le but est alors d'apprécier l'homogénéité des résultats des participants. Ainsi, dans l'ensemble, la question de la quantification globale de la variance d'observation est rarement abordée.

Les données disponibles à travers les réseaux d'observation et de surveillance de l'environnement marin se présentent sous la forme de séries temporelles. A l'évidence, ce ne sont pas des données de plan d'expérience, et ainsi aucun facteur n'est contrôlé, toutes les variabilités s'expriment à travers les mesures réalisées. Les données considérées ne sont généralement pas gaussiennes, pas stationnaires, ni en moyenne ni en variance. Cette non-stationnarité peut être le fait de phénomènes écologiques (e.g. changement globaux, pressions anthropiques) ou artéfactuels (e.g. changement de méthodes, d'agents, de laboratoire). Elles présentent des données manquantes, des données exceptionnelles voire des données fausses. L'ensemble de ces caractéristiques rendent ces séries temporelles particulièrement délicates à traiter. Cependant, les modèles linéaires dynamiques (*i.e.* *Dynamic Linear Models*, DLM)^a permettent de traiter les séries temporelles non-stationnaires comportant des données manquantes et peuvent prendre en compte les changements intervenant dans les séries via des « interventions » (West and Harrison, 1997). De ce fait, ils constituent une approche particulièrement bien adaptée à l'analyse des séries temporelles environnementales marines. Par ailleurs, la variance non expliquée par le modèle fait partie des paramètres estimés dans le processus d'ajustement d'un DLM à une série temporelle. Cette variance peut s'apparenter à une variance d'observation. De fait, la combinaison des approches bayésienne et « espace d'état » a déjà été identifiée comme à même d'incorporer et évaluer les sources d'incertitudes (Harwood and Stokes, 2003) et utilisée comme telle (Hefley et al., 2013). Ainsi notre proposition consiste à ajuster des DLM à des séries temporelles environnementales marines et à considérer les variances d'observations estimées de manière à évaluer la vraisemblance de leurs amplitudes au titre de variabilité globale du processus d'acquisition de la donnée. Cette action Aquaref s'inscrivant dans le contexte de l'hydrobiologie, l'abondance phytoplanctonique va être analysée. La concentration en chlorophylle *a*, bien que relevant du contexte « chimie » d'Aquaref, est également retenue au titre d'indicateur de la biomasse.

2. DONNÉES ET MÉTHODES

2.1 RÉSEAU, LIEUX, PÉRIODE, MÉTHODES DE MESURES

Les données utilisées sont celles du Réseau d'observation et de surveillance du phytoplancton marin et des phycotoxines (REPHY). Les premières observations de ce réseau datent de 1987.

La zone atelier choisie est celle sous la responsabilité du laboratoire environnement ressource (LER) d'Arcachon. Elle couvre l'aire géographique allant de l'estuaire de la Gironde jusqu'au pays Basque. Plus précisément, ce sont les lieux « Arcachon-Bouée 7 » et « Teychan bis » (*cf.* Figure 1). Ces choix se justifient par la longueur des séries temporelles, la régularité de la fréquence de prélèvement, l'absence d'interruption de mesure, seulement deux agents successivement en charge des comptages phytoplanctoniques, changement au premier comptage de 2012, la présence de l'un d'eux sur le site Ifremer de Nantes, la qualité connue et reconnue des identifications et comptages phytoplanctoniques et le contraste environnemental et hydrologique attendu sur les deux lieux de surveillance.

a. Ce type de modèle est également référencé sous le nom de modèle espace d'état (*State space models*) en association avec le filtre de Kalman (*Kalman Filter*). Voir par exemple Harvey et al. (1998).



Figure 1 : Positionnement géographique des lieux de surveillance « Teychan bis » et « Arcachon-Bouée 7 ».

Sur le lieu « Arcachon-Bouée 7 », les prélèvements ont toujours été effectués à la profondeur qualitative « Surface (0-1m) ». En revanche à « Teychan bis », jusques et y compris en 2007, la profondeur qualitative des prélèvements était « Mi-profondeur » avant de s'effectuer à « Surface (0-1m) ». « Mi-profondeur » correspond à des prélèvements réalisés à 3 mètres de profondeur. L'absence de stratification dans le bassin (Neaud-Masson, comm. pers.) permet de supposer une homogénéisation sur la colonne d'eau et ainsi de traiter les deux profondeurs comme identiques. Au demeurant, l'analyse de la série temporelle devrait permettre d'infirmer ou confirmer cette hypothèse.

La période considérée s'étend de 2003 à 2014, bornes incluses. La limitation à 2003 pour toutes les séries temporelles s'est imposée en premier lieu parce que c'est l'année à partir de laquelle des flores totales ont commencé à être réalisées sur le lieu « Arcachon-Bouée 7 » et en second lieu de manière à comparer les séries temporelles « toutes choses égales par ailleurs ». Plus précisément, cette limitation permet de comparer des séries temporelles d'une fenêtre présentant exactement les mêmes événements exogènes, le cas échéant susceptibles d'influencer la variabilité. Tous les prélèvements sont effectués au plus à deux heures de la marée haute, par excès ou par défaut, conformément au protocole REPHY.

Les mesures de chlorophylle *a* ont été effectuées par fluorimétrie (Neveux and Panouse, 1987). La méthode mise à jour (Aminot and Kérouel, 2004) a été utilisée à partir de décembre 2009. Les flores totales sont effectuées selon la méthode d'Utermöhl (Neaud-Masson, 2015; Utermöhl, 1958).

2.2 INFORMATIONS EXOGÈNES, DONNÉES EXTRAITES ET PRÉPARATION

Outre l'expertise biologique, écologique et terrain, les informations exogènes utilisées ici sont contenues d'une part dans les Bulletins de la surveillance du laboratoire

environnement ressource d'Arcachon, édition 2004 à 2015 (*i.e.* surveillance effectuée de 2003 à 2014) et d'autre part dans les bilans climatiques de Météo-France^b.

Les données ont été extraites de la base Quadrige le 24 septembre 2015. La fenêtre temporelle de l'extraction s'étend du 01/01/1900 au 31/12/2014. Pour la chlorophylle *a*, les programmes de mesures sélectionnées sont RNOHYD, REPHY, ARCHYD, SRN, RHLN, RSLHYD et le paramètre est CHLOROA. Pour les comptages phytoplanctoniques les programmes de mesures sont REPHY, SRN, RHLN et le paramètre est FLORTOT.

Les données déjà qualifiées comme fausses ou douteuses ont été écartées. La profondeur qualitative de prélèvement a été ramenée à une valeur unique, la fenêtre temporelle recadrée sur la période 2003-2014, bornes incluses.

Le traitement du paramètre FLORTOT nécessite plusieurs opérations supplémentaires. Pour de multiples raisons évoquées par Hernández-Fariñas (2015), les données phytoplanctoniques acquises comportent une grande variété de niveaux taxinomiques identifiés et taxons virtuels^c. Le regroupement des taxons dans des unités taxinomiques (UT) diminue l'effet de ces variations. Les UT utilisées (*cf.* annexe 1) ici sont celles issues de l'expertise menée par Hernández-Fariñas, marginalement mises à jour pour tenir compte des données les plus récemment acquises. Les ciliés qui sont parfois identifiés dans les comptages mais qui ne sont pas du phytoplancton sont exclus du traitement. L'abondance totale est définie par la somme des abondances des UT appartenant aux phylums *Ochrophyta* (*i.e.* diatomées) et *Myzozoa* (*i.e.* dinoflagellés), les UT n'appartenant pas à ces deux phylums constituant un ensemble marginal et hétéroclite.

Une contrainte de l'approche série temporelle réside dans la nécessité de déterminer l'unité temporelle de base de l'analyse. Ainsi, en tenant compte de la stratégie de prélèvement, de sa fréquence théorique d'observation et de son application effective à travers les données, il s'agit de déterminer la taille de l'unité de temps (*e.g.* jour, semaine, quinzaine, mois) pour laquelle on attend au plus une observation. Du choix de cette unité s'ensuit deux conséquences 1) des unités temporelles devront être créées avec comme observation une donnée manquante dans le cas où elles ne contiennent pas une date de prélèvement, 2) les unités temporelles présentant plus d'une observation doivent être ramenées à une observation en utilisant un opérateur d'agrégation (*e.g.* médiane, maximum, minimum, moyenne). Opérationnellement, un processus automatique détermine pour chaque série son unité temporelle de base minimisant le nombre d'unités générées avec une donnée manquante et celui pour lequel une agrégation temporelle a été nécessaire, le choix de l'opérateur d'agrégation étant la médiane.

Afin de stabiliser les variances dans le temps et ainsi respecter l'hypothèse d'homoscédasticité du modèle, les données sont transformées. Pour la chlorophylle *a*, c'est le logarithme des concentrations qui est modélisé. Pour les comptages cellulaires, c'est le logarithme décimal.

b. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques>

c. Un taxon virtuel est composé d'un ensemble d'espèces, voire de genres, qui ne peuvent être distingués lors de la détermination mais sont plus restreints que le niveau taxinomique supérieur.

2.3 MÉTHODES

Modèle

Un DLM peut contenir différents composants tels qu'une tendance, un terme saisonnier ou certaines variables indépendantes et leurs coefficients. La forme générale du modèle utilisé ici comporte deux parties : une tendance linéaire locale, sous la forme d'un DLM de série temporelle (TSDLM) d'ordre 2, et une composante saisonnière. Les données sont décrites par l'équation d'observation :

$$Y_t = \mu_t + \mathbf{F} \mathbf{S}_t + v_t, \quad v_t \sim N(0, V)$$

avec

- Y_t , observation
- μ_t , niveau moyen ou tendance
- $\mathbf{F} \mathbf{S}_t$, saisonnalité
- v_t , terme d'erreur ou innovation
- V , variance d'observation

Ici et par la suite, les caractères gras représentent des vecteurs et des matrices. Le signal observé est décomposé en un niveau moyen et une saisonnalité. Le vecteur \mathbf{F} dépend de la forme, factorielle ou trigonométrique, et l'unité de temps de la saisonnalité. Enfin, un terme d'erreur v_t , distribué selon une loi normale de moyenne 0 et de variance V est ajouté pour représenter le bruit venant de toute la variabilité du processus d'acquisition des données du prélèvement, incluant la variabilité de l'environnement, à la saisie de données et à des facteurs qui ne sont pas pris en compte dans le modèle. La variance V est la variance d'observation qui fait l'objet de notre attention dans cette étude. La somme du niveau moyen et de la saisonnalité représente un processus inobservable sous-jacent. Ces composants sont indicés par le temps, ce qui souligne qu'ils peuvent évoluer avec le temps. L'équation d'évolution décrit cette évolution :

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{\mu,t}, & \omega_{\mu,t} &\sim N(0,0) \\ \beta_t &= \beta_{t-1} + \omega_{\beta,t}, & \omega_{\beta,t} &\sim N(0, W_{\beta,t}) \\ \mathbf{S}_t &= \mathbf{G} \mathbf{S}_{t-1} + \omega_{S,t}, & \omega_{S,t} &\sim N(0, W_{S,t}) \end{aligned}$$

A l'instant t , le niveau moyen est égal à sa valeur au temps $t-1$ additionné à β_{t-1} . Cela signifie aussi que $\mu_t - \mu_{t-1} = \beta_{t-1}$, *i.e.* la différence de niveau moyen entre deux unités de temps, *i.e.* β_{t-1} est la pente. En résumé, le niveau moyen au temps t est égal au niveau moyen au temps $t-1$ additionné à la pente au temps $t-1$. En outre, un terme d'erreur $\omega_{\mu,t}$ est ajouté. Comme l'équation d'évolution décrit le processus sous-jacent réel et inobservable, ce terme d'erreur est parfois appelé l'innovation, en tant que source de changements. Mais ici, sa variance est égale à zéro, ce qui induit que la seule source de variation dans le niveau moyen est la pente. Ce cas particulier de modèle est appelé *Integrated Random Walk*. Les deux dernières équations décrivent, respectivement, les évolutions de la pente et de la saisonnalité égaux à ceux de $t-1$ avec des termes d'innovation, $\omega_{\beta,t}$ et $\omega_{S,t}$ de variance non nulles. La matrice \mathbf{G} dépend de la forme et de l'unité de temps de la saisonnalité. L'équation d'évolution peut prendre une forme matricielle dans laquelle les termes d'erreurs constituent un vecteur multinormale de moyenne 0 et de variance \mathbf{W}_t : cette dernière est appelée variance d'évolution.

Le TSDLM d'ordre 2 a été choisi pour sa capacité à prendre en charge les tendances constantes, linéaires et quadratique ce qui, *a priori*, semble suffisant pour capturer tout type d'évolution dans les séries chronologiques environnementales. Comme indiqué précédemment, la saisonnalité peut être spécifiée sous deux formes : factorielle et trigonométrique. Dans le premier cas, un paramètre est estimé pour chaque unité de

temps (e.g. mois). Si la saisonnalité est supposée être constante dans le temps (c.-à-d. variance d'innovation égale à 0) alors les mêmes estimations sont utilisées pour chaque unité de temps au fil des ans. A l'inverse, le paramètre de saison peut être estimé à chaque unité de temps reflétant une saisonnalité variable dans le temps. Si la saisonnalité est modélisée sous forme trigonométrique, en substance, une phase et une amplitude sont estimées qui peuvent être constantes ou variables dans le temps en fonction de la spécification de $W_{s,t}$. Si la saisonnalité est soupçonnée être bi-modale, c'est-à-dire présenter deux épisodes de valeurs élevées, deux harmoniques doivent être précisées de sorte que deux phases et deux amplitudes soient estimées. Indiquer deux harmoniques n'induit pas nécessairement un modèle présentant deux modes mais permet de modéliser un système présentant au plus deux modes.

Dans le choix de la forme de la saisonnalité, l'approche trigonométrique, impliquant une évolution sinusoidale, peut être vue comme une hypothèse trop forte et ainsi la forme factorielle pourrait être toujours préférée. Toutefois, le principe de parcimonie incline à choisir, à performance égale, les modèles présentant le moins de paramètres. Or dans la forme factorielle, le nombre de paramètres de saisonnalité est égal au nombre de « saisons » moins 1. Pour les unités allant du trimestre au jour, le nombre de paramètres augmente de 3 à 364 pour la forme factorielle tandis que la forme trigonométrique ne nécessite que deux paramètres par harmonique. Au regard d'une fréquence de prélèvement théorique bimensuelle et de l'observation habituelle d'un bloom printanier et d'un bloom automnal, en respect du principe de parcimonie une forme trigonométrique avec deux harmoniques a été choisie.

Lorsque les composantes du modèle ont été choisies, les paramètres du modèle, c'est-à-dire les variances d'observations et d'évolutions, peuvent être estimées par la méthode du maximum de vraisemblance. Pour ce processus d'optimisation, les valeurs initiales ont été choisies égales à la variance de la série temporelle modélisée. Les variances d'observation et d'innovation jouent un rôle central en estimant la part de signal et la part de bruit dans les observations. Le ratio signal sur bruit détermine la mémoire du processus et la quantité d'informations contenue dans les observations utilisées pour le filtrage et lissage des observations. Ces dernières étapes sont prises en charge par un filtre de Kalman. Enfin, comme le processus est défini de façon récursive, avec la première observation à $t=1$, des valeurs *a priori* à $t=0$ pour le niveau moyen, la pente et la saisonnalité et leurs variances doivent être spécifiées. Des valeurs très peu informatives sont choisies : toutes les moyennes sont mises à 0 et toutes les variances à 10^7 . Ces valeurs sont celles utilisées par défaut (Petris et al., 2009).

Interventions

Une intervention est le nom donné à une modification d'un paramètre du modèle afin de prendre en compte une information exogène. Ici, seuls les changements de niveau moyen et la prise en compte des valeurs exceptionnelles sont considérés. Précédemment, il a été souligné que la pente est le seul facteur de l'évolution de la tendance, car la variance d'évolution du niveau moyen est fixée à 0. Quand un changement dans le niveau moyen est suspecté à l'instant t , une variance d'évolution du niveau moyen non nulle est spécifiée pour cette unité de temps et estimée par maximum de vraisemblance. Pour ce processus d'optimisation, la valeur initiale est la variance de la série temporelle modélisée. Il peut y avoir autant de variances d'évolution non nulles du niveau moyen que de changements de niveau soupçonnés. Ces spécifications n'induisent pas nécessairement un modèle montrant des changements de niveau mais les probabilités de changement seront évaluées et, si certains d'entre eux sont vraisemblables d'un point de vue numérique, le modèle présentera des changements de niveau. Les experts restent juges ultimes de la pertinence de ces interventions.

Les valeurs exceptionnelles sont définies comme des mesures avec des variances d'observation inhabituellement élevées, quelle que soit leur source (e.g. événements inhabituels). Si une valeur exceptionnelle est suspectée à l'instant t , alors un paramètre d'augmentation de la variance d'observation est ajouté au modèle. Il s'agit d'une quantité supérieure ou égale à 1 qui intervient comme multiplicateur de la variance d'observation de routine. Chaque valeur exceptionnelle peut être traitée de cette façon. Comme auparavant, la pertinence numérique de ces spécifications est évaluée par maximum de vraisemblance, la valeur initiale pour le processus d'optimisation étant 1, et les experts ont le dernier mot quand à l'identification effective des valeurs exceptionnelles.

Identification des valeurs exceptionnelles et des changements dans la tendance

Trois possibilités étaient envisageables pour l'identification des points de changement et des valeurs exceptionnelles. La première était d'avoir recours exclusivement à des approches statistiques. De cette manière, les résultats présenteraient un caractère d'optimalité d'un point de vue numérique mais le cas échéant aux dépens de l'interprétation écologique. La seconde option était de se fonder uniquement sur l'information exogène. Ainsi chaque changement est motivé par une causalité préexistante ; cependant il n'existe pas de série pour laquelle une telle information existe de manière exhaustive, qui au demeurant, rendrait l'analyse inutile. La dernière approche consiste à s'appuyer sur des méthodes statistiques et la connaissance des experts. C'est celle qui est mise en œuvre ici.

L'approche statistique utilisée pour la détection des changements et valeurs exceptionnelles s'appuie sur la définition des valeurs exceptionnelles dans la représentation boîte-à-moustaches (*Box-and-whiskers plot*). Dans cet outil, les valeurs exceptionnelles d'une loi normale centrée-réduite sont les valeurs supérieures à 2.7 ou inférieures à -2.7, ce qui correspond, respectivement, aux 0.35 % valeurs les plus élevées et les 0.35 % valeurs les plus faibles pour un total de 0.7 % de la distribution dans son ensemble. Ces valeurs seuils, -2.7 et 2.7, sont utilisées en conjonction avec les résultats d'un modèle. Pour les valeurs exceptionnelles, les erreurs standardisées sont examinées. Il s'agit des écarts entre les observations et les valeurs lissées du modèle, divisées par la racine carrée des variances des valeurs lissées. La distribution des erreurs standardisées est supposée être une loi normale centrée-réduite. De ce fait, les valeurs supérieures à 2.7 ou inférieures à -2.7 correspondent à des observations potentiellement exceptionnelles et ainsi candidates à un traitement approprié.

La même démarche est mise en œuvre pour les changements de niveau. Les valeurs examinées sont appelées résidus auxiliaires (*auxiliary residuals*) (Harvey et al., 1998). Pour les changements de niveaux, il faut considérer les résidus auxiliaires du niveau moyen c'est à dire les valeurs lissées du terme d'erreur du niveau moyen $\omega_{\mu,t}$. Dans le cas particulier de l'*Integrated Random Walk* utilisé ici, comme la variance de son terme d'erreur est nulle, ses valeurs lissées sont également nulles et ainsi non-informatives sur les changements de niveau. Cependant, l'idée des *auxiliary residuals* est d'examiner les innovations lissées, c'est à dire les changements d'une unité de temps à l'autre, c'est-à-dire les différences premières du niveau moyen, $\mu_t - \mu_{t-1}$. Or comme $\mu_t - \mu_{t-1} = \beta_{t-1}$, les valeurs lissées de β_{t-1} standardisées par leurs variances lissées et centrées sur leur moyenne portent une information similaire. Par convention et pour la clarté du propos à venir, ces valeurs sont désignées ci-après par l'expression « résidus auxiliaires du niveau ». Elles peuvent être comparées aux valeurs seuils 2.7 et -2.7 au-delà desquelles s'impose le diagnostic de valeurs de pentes exceptionnelles, c'est-à-dire que l'évolution des données pousse le modèle dans les limites de son adaptation, et suggère un changement dans le niveau. Contrairement au cas des valeurs exceptionnelles pour lequel une erreur standardisée est associée à une mesure, les résidus auxiliaires exceptionnels de la pente constituent souvent des séquences : la mesure pour laquelle il convient de suggérer un changement de

niveau n'est pas unique. L'expérience a montré qu'utiliser la valeur la plus élevée n'est pas toujours pertinent dans la mesure où l'événement induisant l'adaptation exceptionnelle peut avoir eu lieu avant ou après la valeur la plus élevée, et peut même s'être produit avant ou après la séquence de résidus auxiliaires exceptionnels de la pente. Compte tenu de cela, la procédure opérationnelle d'identification de la mesure candidate pour une suggestion de changement de niveau est définie comme suit. Lorsque les résidus auxiliaires de la pente constituent une séquence exceptionnelle (*i.e.* supérieures à 2.7 ou inférieures à -2.7), la séquence est étendue aux valeurs voisines supérieures à 2 ou inférieures à -2^d et, dans cette nouvelle séquence, la mesure présentant l'erreur standardisée la plus élevée en valeur absolue est utilisée pour porter la suggestion de changement de niveau.

Stratégie d'analyse semi-automatisée

L'ensemble des éléments de méthode décrits ci-avant ne définissent pas une approche dans laquelle l'estimation des paramètres et l'identification des changements et valeurs exceptionnelles sont réalisées conjointement. Par voie de conséquence, une stratégie d'analyse doit être définie. Comme l'identification statistique des valeurs exceptionnelles et des changements de niveau s'appuie sur les résultats d'un modèle, il y a nécessité d'ajuster les paramètres du modèle une première fois. Les changements et *outliers* suggérés définissent un second modèle, *etc.* Ainsi, potentiellement, le premier modèle constitue l'amorce d'une procédure itérative conduisant à d'autres modèles, le cas échéant de manière infinie. Cet écueil est évité du fait de l'identification des *outliers* et changements de niveaux avec les seuils (-2.7, 2.7) correspondant à 0.7 % des distributions. De cette manière, et par expérience, il n'est pas rare que le premier modèle ne présente pas de suggestion d'intervention tandis que, le plus souvent, la stabilité est obtenue au second modèle ou au troisième modèle. Ainsi, le nombre par défaut d'itération a été limité à trois. Plus rarement le troisième modèle suggère encore des changements. Il est alors possible de les prendre en compte « manuellement ».

Diagnostic des modèles

Les deux hypothèses à vérifier sont la normalité et l'indépendance des résidus standardisés. La première est traditionnellement appréciée à l'aide d'un diagramme quantile-quantile ou Q-Q plot : les quantiles théoriques de la loi normale centrée-réduite sont portés en abscisse et les valeurs correspondantes des quantiles de la distribution observée centrée-réduite en ordonnée. Lorsque les points ne présentent pas de déviation majeure à la première bissectrice, alors la distribution observée suit probablement une loi normale. Cette approche peut être complétée par un test statistique formel. Le test de Shapiro-Wilk est considéré comme le plus puissant, c'est à dire le moins enclin à conclure à la normalité en son absence. Toutefois, sa sensibilité aux déviations dans les queues de distribution et son utilisation pour des échantillons de taille inférieure à 50 sont souvent relevés. Le test de Kolmogorov-Smirnov est réputé moins puissant mais également moins sensible aux faibles écarts à la normalité. L'approche graphique et ces deux tests sont utilisés ici pour apprécier la normalité des résidus.

L'indépendance des résidus est fréquemment évaluée à l'aide du test de Ljung-Box. En présence d'une série temporelle comportant des données manquantes, l'implémentation de ce test dans R^e (R Development Core Team, 2008) produit un résultat bien qu'officiellement, il ne les prenne pas en charge : les calculs élémentaires impliquant une donnée manquante sont ignorés. Si cette approche peut être d'un impact minimal lorsque le pourcentage de données manquantes est très réduit, intuitivement, il sera d'autant plus important que ces proportions seront plus élevées. Or, les séries temporelles de cette étude peuvent comporter plus de 50 % de données manquante. Stoffer et Toloï (1992) ont

d. Les valeurs 2 et -2 d'une loi normale centrée réduite correspondent aux quantiles 2.5 % et 97.5 % classiquement utilisés en statistique inférentielle.

e. R est un environnement logiciel libre pour le calcul statistique et la production de graphiques.

proposé une version modifiée du Ljung-Box pour prendre en charge les données manquantes. C'est cette version qui a été programmée et est ici utilisée.

Pertinence des interventions, prolongement de l'analyse et intégration de l'expertise

La qualité de l'ajustement d'un modèle est appréciée à l'aide de la log-vraisemblance (*i.e.* *log-likelihood*, LL) qui est d'autant plus élevée que le modèle est numériquement vraisemblable. Le caractère significatif de la différence de log-vraisemblance entre deux modèles emboîtés peut être testé. Deux modèles sont emboîtés si l'ensemble des paramètres de l'un est inclus dans l'ensemble des paramètres de l'autre. Dans ce cas, la différence des log-vraisemblances multipliée par deux suit une loi du Chi² dont le nombre de degrés de libertés est la différence du nombre de paramètres entre les deux modèles. Par ailleurs, les comparaisons entre modèles sont réalisées à l'aide du critère d'information d'Akaike (*i.e.* *Akaike information criterion*, AIC) qui pénalise les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie : le meilleur modèle est celui avec le plus faible AIC. Ces deux approches peuvent être utilisées conjointement dans la mesure où un nouveau modèle peut présenter un gain significatif de log-vraisemblance et une augmentation du critère d'Akaike.

Les mesures précédentes peuvent être utilisées pour évaluer rétrospectivement les modèles de l'analyse semi-automatisée. Le cas échéant, cela peut conduire à ne pas identifier l'ultime modèle comme le plus pertinent. Il est également possible d'exclure certaines interventions considérant que, si elles présentent un caractère vraisemblable d'un point de vue numérique, leur contribution est limitée d'un point de vue explicatif. Il est par ailleurs souhaitable, bien que pas indispensable, de pouvoir justifier les interventions conservées dans le modèle. Inversement, une information exogène ne correspondant à aucune des interventions conservées peut être utilisée pour suggérer un nouveau changement.

Finalement, l'examen des résidus standardisés est également une source d'identification de changement dans la série. En effet, comme ils sont supposés être distribués indépendamment et identiquement selon une loi normale centrée réduite, la probabilité pour un résidu d'être positif ou négatif est 0.5. Il s'ensuit que la probabilité d'avoir deux résidus consécutifs de même signe est $0.5^2 = 0.25$, trois résidus consécutifs de même signe $0.5^3 = 0.125$, etc. La probabilité d'observer une séquence de 7 valeurs consécutives du même signe est de 0.0078125 soit 0.78 %. Ainsi en référence à l'outil « Boîte à moustache », une séquence de 7 résidus consécutifs de même signe présente un caractère exceptionnel qui témoigne d'une inadaptation locale du modèle aux données et peut être un indice d'un changement de niveau. Plus les valeurs de la séquence sont éloignées de 0 en valeur absolue, plus la probabilité d'un changement de niveau est élevée. Les changements envisagés peuvent être suggérés au modèle et leurs vraisemblances numériques évaluées. Si les changements suggérés apparaissent pertinents, alors il est très souhaitable de pouvoir les relier explicitement à des informations exogènes.

Représentations et expressions de la variance d'observation

Lorsqu'un modèle est considéré comme final, la question du caractère vraisemblable de sa valeur estimée de la variance d'observation peut être posée. Mais d'une part, la variance d'une distribution est une quantité abstraite, d'autre part, les modélisations réalisées concernent une transformée logarithmique de la variable d'intérêt. Il y a donc lieu d'effectuer un travail autour de l'expression et de la représentation de la variance d'observation.

A chaque instant t de la série temporelle, le DLM fournit la moyenne et la variance de la distribution normale de la variable d'intérêt modélisée. Pour la chlorophylle a , cette dernière est le logarithme des concentrations. L'exponentielle de la moyenne du logarithme de la concentration n'est pas la moyenne de la concentration mais la médiane de la concentration. Les moyennes et les variances des distributions log-normales des concentrations peuvent être obtenues à partir des formules suivantes :

$$\begin{aligned} \text{Si } Y = \ln(X) \text{ alors} \\ E(X) &= e^{E(Y) + V(Y)/2} \\ V(X) &= e^{2E(Y)}(e^{2V(Y)} - e^{V(Y)}) \end{aligned}$$

avec $E(X)$ et $V(X)$, respectivement, la moyenne et la variance de X . Le graphique de ces variances en fonction des moyennes est une courbe quadratique typique des lois log-normales. Ce résultat est induit par la transformation logarithmique utilisée. Pour la même raison, la figure des logarithmes décimaux des variances en fonction des logarithmes décimaux des moyennes représente une droite dont la pente est très proche de 2. Ce résultat mécanique mis à part, l'ordonnée à l'origine est la seconde caractéristique de la droite. Elle peut être comparée à une valeur de référence. Brown et al. (1998) ont calculé les paramètres d'une telle droite à partir de mesures de chlorophylle a réalisées dans 209 lacs de Floride durant au moins un an à raison de 11 à 13 mesures par an. C'est une référence à laquelle on peut comparer les droites obtenues avec les résultats des modèles. Pour les comptages phytoplanctoniques, la variable d'intérêt est le logarithme décimal des abondances. Comme c'est l'unité adaptée de représentation de ces quantités, les considérations précédentes concernant la chlorophylle a ne s'appliquent pas.

Une seconde représentation pour appréhender la variabilité pour la chlorophylle a consiste à porter en abscisse les médianes des distributions des concentrations (*i.e.* les exponentielles des moyennes des logarithmes des concentrations), en ordonnée les valeurs observées et d'ajouter les enveloppes de confiance des observations et de la médiane. La surface définie par l'enveloppe de confiance des observations est une expression de la variabilité totale des mesures. La surface délimitée par l'enveloppe de confiance de la médiane est une expression de la variabilité liée à l'évolution de la somme de la tendance et de la saisonnalité. Le même type de représentation peut être réalisé pour les logarithmes décimaux des abondances phytoplanctoniques, les moyennes des distributions se substituant aux médianes.

Ainsi dans la variabilité totale, une part concerne l'évolution structurelle, *i.e.* la somme de la saisonnalité et de la tendance. Cette part peut être exprimée par le rapport de l'étendue de l'intervalle de confiance de la médiane de la concentration en chlorophylle a à celui des observations, rapport qui peut également prendre la forme d'un pourcentage. Le pourcentage complémentaire concerne la surface entre l'enveloppe de confiance des médianes et celui des observations. Dans le cadre de la question sur les incertitudes, c'est cette dernière qui est l'objet d'intérêt et c'est celle qui est estimée par la variance d'observation vue précédemment. Ces pourcentages constituent une autre forme de représentation des variabilités.

Une dernière expression de la variabilité consiste à définir la précision en pourcentage de la mesure effectuée. Typiquement, la mesure X réalisée à la précision P % signifie que l'intervalle de confiance à 95 % de la mesure X est $[X - X.P/100 ; X + X.P/100]$ (Aminot and Kérouel, 2004). Ainsi, l'intervalle de confiance à 95 % d'une valeur permet de calculer les distances aux bornes en pourcentages de la valeur. Ces pourcentages à la borne supérieure et à la borne inférieure sont les mêmes pour les distributions symétriques (*e.g.* gaussiennes) mais sont différents pour les distributions asymétriques (*e.g.* log-normales). Ces calculs peuvent être réalisés avec les intervalles de confiance des observations. Mais comme précédemment, c'est l'écart entre l'enveloppe de confiance des observations et

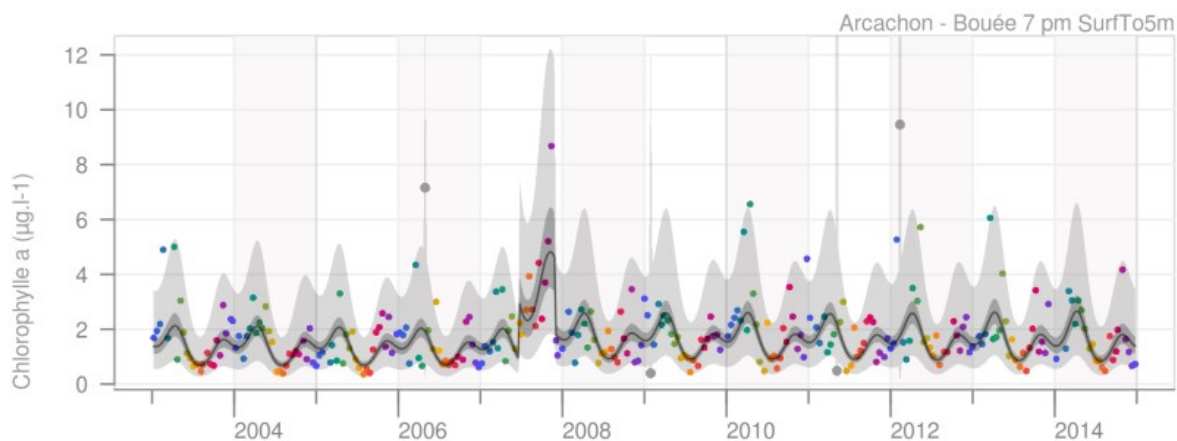


Figure 2 : Résultat du modèle linéaire dynamique expertisé pour la concentration en chlorophylle a au lieu de surveillance « Arcachon-Bouée 7 ». Les disques représentent les observations. La couleur des disques est relative à la saison des prélèvements : les teintes bleues correspondent à l'hiver, les vertes au printemps, les jaunes-orangées à l'été et du rouge au violet à l'automne. La ligne continue représente la valeur du modèle, c'est à dire la médiane des distributions modélisée, c'est à dire le niveau moyen additionné à la saisonnalité. La surface gris foncée est l'enveloppe de confiance à 95 % du modèle. La surface gris clair est l'enveloppe à 95 % des observations. Les disques gris sont les observations traitées comme exceptionnelles.

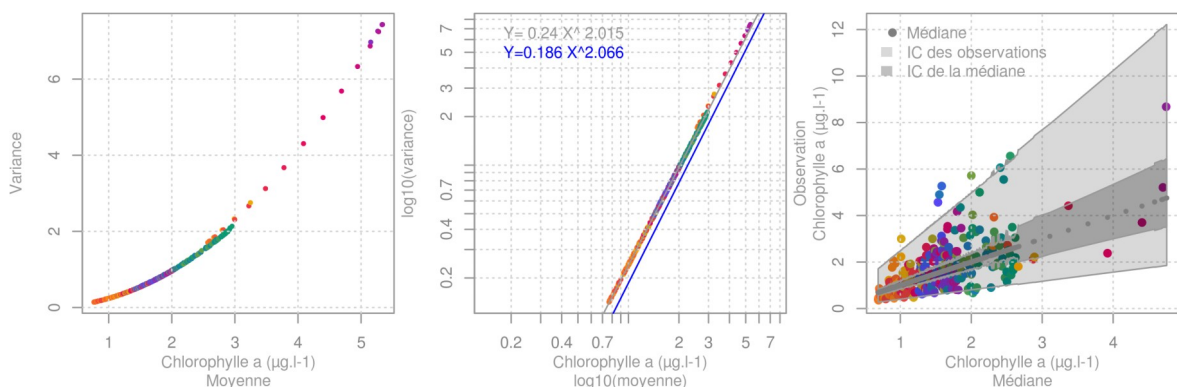


Figure 3 : Représentations de la variance totale du modèle linéaire expertisé pour la concentration en chlorophylle a au lieu de surveillance « Arcachon-Bouée 7 ». Les disques représentent les observations. Le code couleur est identique à celui décrit pour la figure 2. A gauche : variance exprimée en fonction de la moyenne de concentration. Au centre : même chose en échelle logarithmique ; la droite grise est la droite des moindres carrés pour les observations, la droite bleue est la droite de référence de Brown et al. (1998). A droite : les observations sont représentées en fonction de la médiane de leurs distributions, conjointement aux limites à 95 % du modèle et des observations.

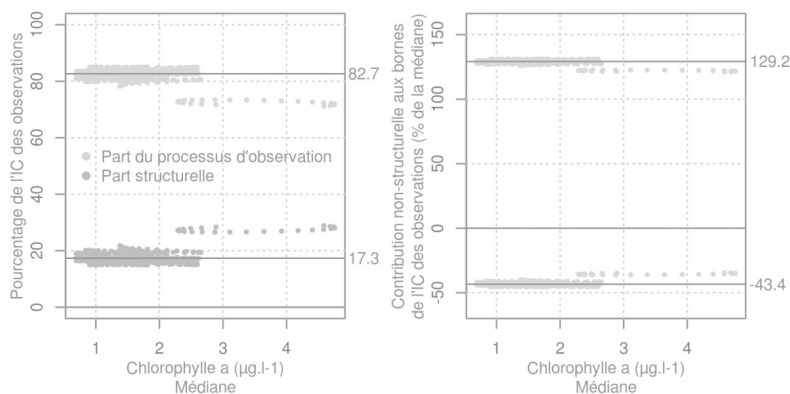


Figure 4 : Représentations et expressions de la variance d'observation pour la concentration en chlorophylle a au lieu de surveillance « Arcachon-Bouée 7 ». A gauche : parts du processus d'observation et de l'évolution structurelle (i.e. tendance et saisonnalité) dans l'intervalle de confiance des observations. A droite : précision (i.e. bornes de l'intervalle de confiance des observations exprimées en pourcentage de la médiane) attribuable au processus d'observation. Les résultats indiqués dans les marges droites des graphiques sont des moyennes.

celle de la médiane qui témoigne de la variabilité du processus d'observation. Rapporté à la médiane, il constitue la part non-structurale de la précision.

Les approches développées dans les deux derniers paragraphes sont applicables aussi bien à la chlorophylle *a* en $\mu\text{g/L}$ qu'à l'abondance en \log_{10} cell./L. Toutefois, dans ce dernier cas, les représentations seront également faites en cell./L, unité plus familière aux experts des comptages phytoplanctoniques.

3. RÉSULTATS ET EXPERTISES

L'unité temporelle de base a été identifiée comme étant la semaine pour toutes les séries traitées. Pour chacune d'elles, l'approche précédemment décrite a été appliquée. Les étapes successives des expertises des modèles sont données en annexe 2. Enfin, les résultats ont été validés d'un point de vue biologique et écologique, par dire d'experts.

3.1 CHLOROPHYLLE *a*

La figure 2 montre le résultat du modèle expertisé pour le lieu de surveillance « Arcachon-bouée 7 ». Le modèle inclut des changements de niveau les 26/06/2007 et 19/11/2007. Les observations des 04/05/2006, 03/02/2009, 10/05/2011 et 16/02/2012 sont traitées comme exceptionnelles. Les résidus satisfont aux exigences de normalité et d'indépendance. La variance d'observation estimée est de 0.208 log chlorophylle *a* $\mu\text{g/L}$. La figure 3 montre les représentations de la variance totale. Comme annoncé précédemment, la relation entre la moyenne et la variance des concentrations du modèle est quadratique. En échelle logarithmique, la droite des moindres carrés des observations est très proche de la droite de référence de Brown et al. (1998). Enfin, l'intervalle de confiance des observations encadre les données dans un cône qui apparaît vraisemblable. La figure 4 montre plus spécifiquement les expressions de la variance d'observation. En moyenne, le processus d'observation contribue pour 83 % à l'intervalle de confiance des observations, la part structurale, *i.e.* la part expliquée par l'évolution du niveau moyen additionné à la saisonnalité constituant les 17 % restants. Exprimé en pourcentage de la médiane, la variabilité attribuable au processus d'observation place la borne inférieure de l'intervalle de confiance des observations à -43 % de la valeur de la médiane et la borne supérieure à +130 %.

La figure 5 montre le résultat du modèle expertisé pour le lieu de surveillance « Teychan bis ». Ce modèle n'inclut aucun changement de niveau. Les observations des 20/04/2006, 11/02/2011 et 21/07/2014 sont traitées comme exceptionnelles. Les résidus satisfont aux exigences de normalité et d'indépendance. La variance d'observation estimée est de 0.176 log chlorophylle *a* $\mu\text{g/L}$. Les représentations de la variance totale de la figure 6 sont très similaires à celles de la figure 3. Les expressions de la variance d'observation de la figure 7 sont du même ordre de grandeur que pour « Arcachon-bouée 7 ». Le processus d'observation contribue pour 78 % à l'intervalle de confiance des observations, la part structurale constituant les 22 % restants. Exprimé en pourcentage de la médiane, la variabilité attribuable au processus d'observation place la borne inférieure de l'intervalle de confiance des observations à -39 % de la valeur de la médiane et la borne supérieure à +111 %.

Un test de rapport des variances de Fisher montre que les variances d'observations estimées sur ces deux lieux de surveillance sont non significativement différentes.

3.2 ABONDANCE PHYTOPLANCTONIQUE

La figure 8 montre le résultat du modèle expertisé pour le lieu de surveillance « Arcachon-Bouée 7 ». Le modèle inclut des changements de niveau les 25/07/2007 et 4/12/2007. Les observations des 7/03/2005, 16/02/2012, 11/10/2013 et 31/10/2014 sont

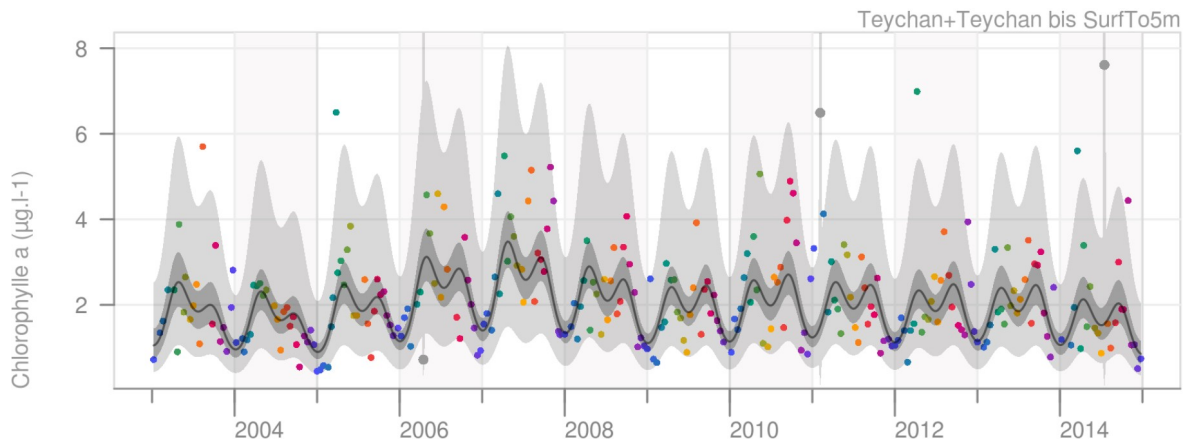


Figure 5 : Résultat du modèle linéaire dynamique expertisé pour la concentration en chlorophylle a au lieu de surveillance « Teychan bis ». cf. Figure 2.

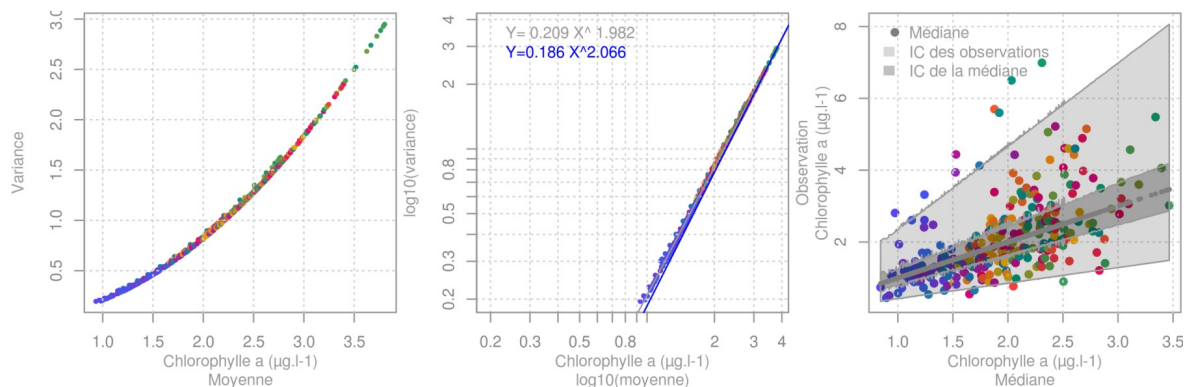


Figure 6 : Représentations de la variance totale du modèle linéaire expertisé pour la concentration en chlorophylle a au lieu de surveillance « Teychan bis ». cf. Figure 3.

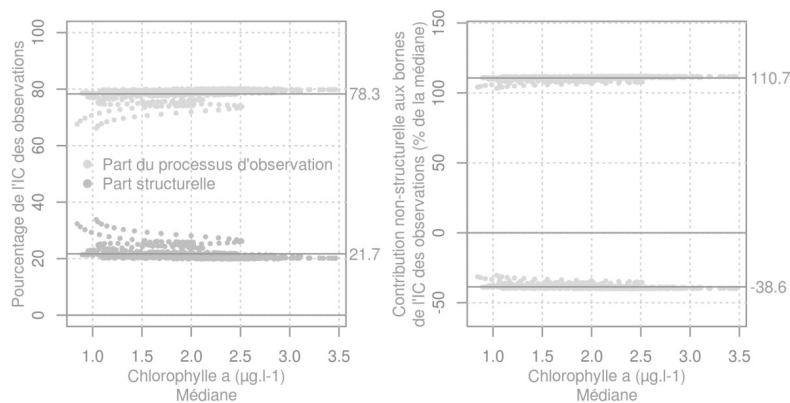


Figure 7 : Représentations et expressions de la variance d'observation pour la concentration en chlorophylle a au lieu de surveillance « Teychan bis ». cf. Figure 4.

traitées comme exceptionnelles. Les résidus satisfont aux exigences de normalité et d'indépendance. La variance d'observation estimée est de 0.204 \log_{10} cell./L. La figure 9 montre que les données sont dans une enveloppe de confiance formant une bande qui apparaît vraisemblable. La part structurelle constitue 21 % de l'intervalle de confiance des observations, les 79 % restants sont attribués au processus d'observation. Exprimé en pourcentage de la moyenne, la variabilité attribuable au processus d'observation place les

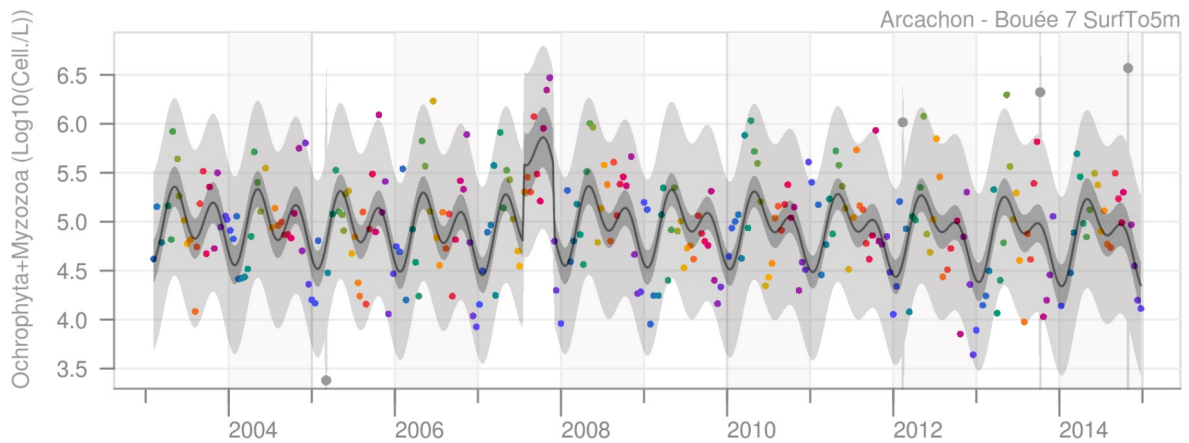


Figure 8 : Résultat du modèle linéaire dynamique expertisé pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochromytha et Myzozoa (i.e. diatomées et dinoflagellés) au lieu de surveillance « Arcachon - Bouée 7 ». cf. Figure 2.

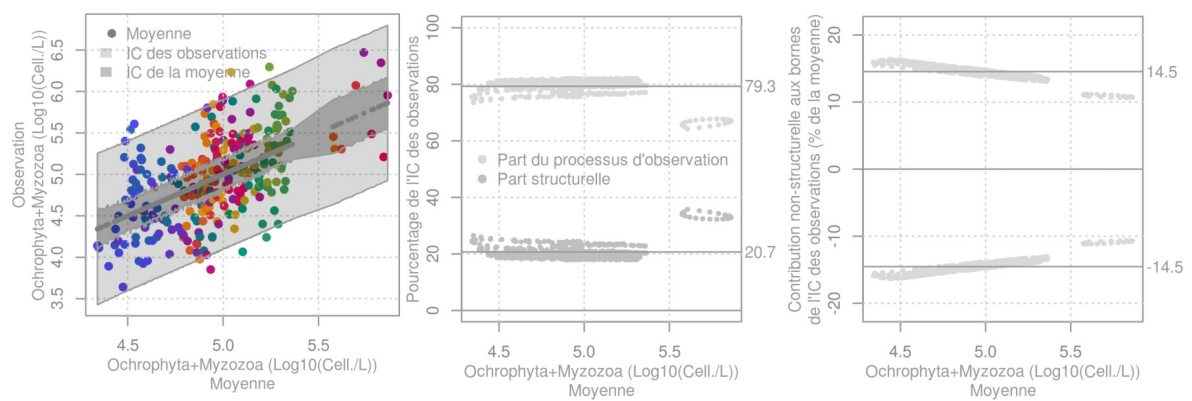


Figure 9 : Représentations de la variance totale du modèle linéaire expertisé (à gauche, cf. Figure 3.) et représentations et expressions de la variance d'observation pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochromytha et Myzozoa (i.e. diatomées et dinoflagellés) au lieu de surveillance « Arcachon - Bouée 7 » (au centre et à droite cf. Figure 4).

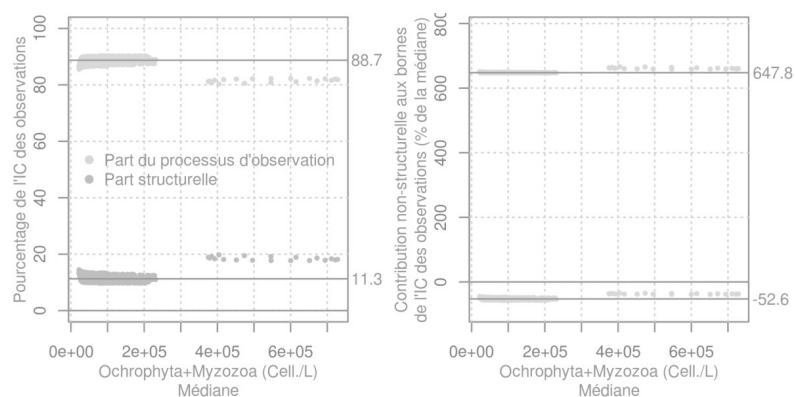


Figure 10 : Représentations et expressions de la variance d'observation pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochromytha et Myzozoa (i.e. diatomées et dinoflagellés) en cell./L au lieu de surveillance « Arcachon - Bouée 7 ». cf. Figure 4.

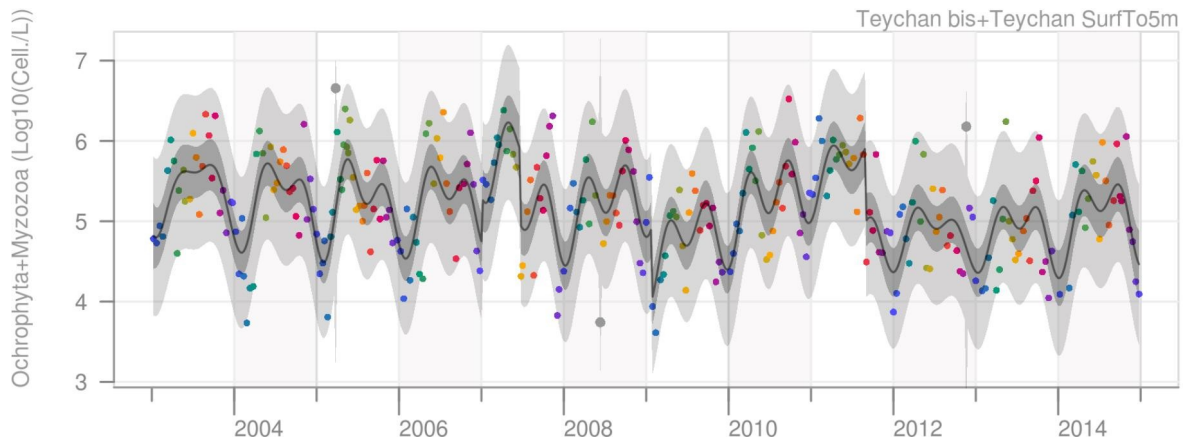


Figure 11 : Résultat du modèle linéaire dynamique expertisé pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochrophyta et Myzozoa (i.e. diatomées et dinoflagellés) au lieu de surveillance « Teychan bis ». cf. Figure 2.

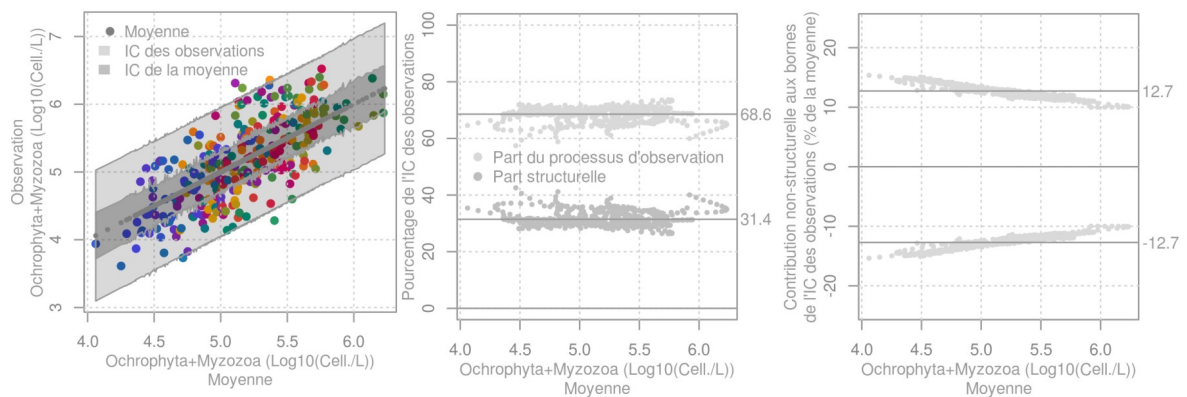


Figure 12 : Représentations de la variance totale du modèle linéaire expertisé (à gauche, cf. Figure 3.) et représentations et expressions de la variance d'observation pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochrophyta et Myzozoa (i.e. diatomées et dinoflagellés) au lieu de surveillance « Teychan bis » (au centre et à droite cf. Figure 4).

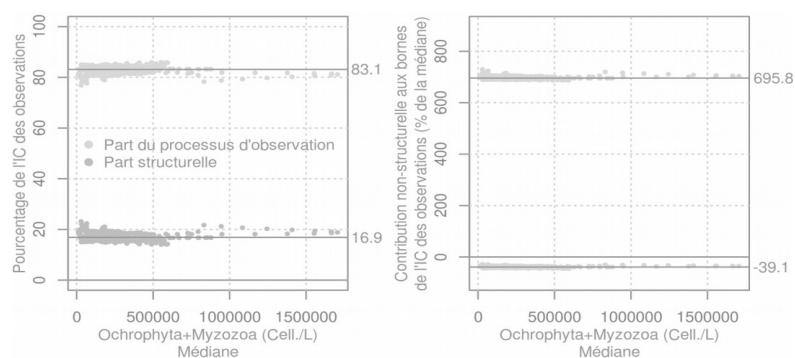


Figure 13 : Représentations et expressions de la variance d'observation pour l'abondance phytoplanctonique estimée par la somme des cellules des phylum Ochrophyta et Myzozoa (i.e. diatomées et dinoflagellés) en cell./L au lieu de surveillance « Teychan bis ». cf. Figure 4.

bornes de l'intervalle de confiance des observations à ± 14.5 % de la valeur de moyenne. La figure 10 exprime ces deux dernières expressions en cell./L. Dans cette échelle linéaire, la part structurelle passe à 11 % et la part attribuable au processus d'observation à 89 %. La variabilité liée au processus d'observation place la borne inférieure de l'intervalle de confiance des observations à -53 % de la valeur de la médiane et la borne supérieure à +648 % de la valeur de la médiane.

La figure 11 montre le résultat du modèle expertisé pour le lieu de surveillance «Teychan bis». Le modèle inclut des changements de niveau les 12/01/2007, 26/06/2007, 3/02/2009 et 6/09/2011. Les observations des 29/03/2005, 12/06/2008 et 22/11/2012 sont traitées comme exceptionnelles. Les résidus satisfont aux exigences de normalité et d'indépendance. La variance d'observation estimée est de $0.212 \log_{10}$ cell./L. La figure 12 montre que les données s'inscrivent dans une enveloppe de confiance formant une bande qui apparaît vraisemblable. La part structurelle constitue 31 % de l'intervalle de confiance des observations ; les 69 % restants sont attribués au processus d'observation. Exprimé en pourcentage de la moyenne, la variabilité attribuable au processus d'observation place les bornes de l'intervalle de confiance des observations à ± 12.7 % de la valeur de moyenne. La figure 13 exprime ces deux dernières expressions en cell./L. Dans cette échelle linéaire, la part structurelle passe à 17 % et la part attribuable au processus d'observation à 83 %. La variabilité liée au processus d'observation place la borne inférieure de l'intervalle de confiance des observations à -39 % de la valeur de la médiane et la borne supérieure à +696 % de la valeur de la médiane.

Un test de rapport des variances de Fisher montre que les variances d'observations estimées sur ces deux lieux de surveillance sont non significativement différentes.

4. DISCUSSION

Les résultats des modèles expertisés ont montré que les séries considérées présentaient des valeurs exceptionnelles et des changements de niveau. L'identification et le traitement approprié des données exceptionnelles sont de première importance. La méthode proposée permet de pointer des données qui sont hors norme relativement à l'ensemble des données mais également hors normes relativement à la saisonnalité. Ces données identifiées, l'expertise a, en premier lieu, pour objet de décider si ce sont des données valides ou fausses. Dans ce dernier cas, la donnée doit être éliminée du traitement. Dans les séries traitées, aucune des données pointées comme exceptionnelles n'a été qualifiée de fausse. En second lieu, lorsque la donnée est considérée comme valide, l'alternative proposée à l'expertise est le traitement comme valeur exceptionnelle ou comme valeur normale. Les décisions prises par les experts ont conduit à ce que toutes les données identifiées comme exceptionnelles soient traitées comme telles. Ce résultat est directement lié à la méthode d'identification des valeurs exceptionnelles, en tant qu'observations appartenant aux 0.7 % des données les plus marginales, par excès ou par défaut. Ainsi les données identifiées sont par définition exceptionnelles, relativement à la modélisation établissant cette identification : la question de l'expertise ne concerne donc plus que la qualification des données en bonnes ou fausses.

De la qualification d'une donnée hors norme comme « exceptionnelle » découle un traitement consistant à estimer pour cette donnée une variance d'observation spécifique supérieure à la variance d'observation standard. Cette approche a plusieurs conséquences. La valeur est conservée dans l'ensemble des données constituant la série temporelle mais son poids est moindre comparativement aux autres données. Ainsi cette observation valide et exceptionnelle a une influence réduite dans l'estimation des autres paramètres du modèle mais également dans le filtrage et le lissage des observations. Au premier chef, c'est l'estimation de la variance d'observation qui est affectée. En tant que donnée hors

norme, la valeur exceptionnelle présente souvent un écart important aux autres valeurs et systématiquement au modèle. Cet écart contribue de manière importante à la variance d'observation. Une fois la donnée traitée comme exceptionnelle, l'écart au modèle possède un poids moindre et ainsi la variance d'observation estimée est réduite, s'approchant d'une estimation « hors valeurs exceptionnelles ». Les autres paramètres du modèle sont également et mécaniquement affectés par le traitement, mais dans une moindre mesure. La diminution de la variance d'observation avec le traitement approprié des données exceptionnelles induit une part de bruit dans la série temporelle moindre que celle initialement estimée. Comme l'information de la série temporelle est, dans la modélisation, partagée entre le bruit lié à l'observation et le signal sous-jacent inobservable (*i.e.* part structurelle correspondant à la somme du niveau moyen et de la saisonnalité), une diminution de la part du bruit entraîne une augmentation de la part signal. Le ratio « signal sur bruit » est également modifié. Or ce ratio contrôle directement l'adaptabilité du modèle : un signal faible induit un modèle à mémoire longue et faible adaptabilité alors qu'un signal fort entraîne un modèle à mémoire courte et forte adaptabilité. Ainsi la diminution de la variance d'observation avec le traitement approprié des données exceptionnelles permet d'une part, de tendre vers une valeur plus réaliste et, d'autre part, d'obtenir un « ratio signal sur bruit » et une adaptabilité en adéquation avec le processus standard, *i.e.* hors données exceptionnelles. Conjointement, la log-vraisemblance augmente, traduisant le meilleur ajustement du modèle aux données tout à la fois sous l'effet du traitement particulier des valeurs exceptionnelles et d'une adaptabilité accordée au processus standard. L'ensemble de ces considérations illustre clairement l'importance d'identifier et de traiter de manière appropriée ces valeurs qui ne sont pas fausses mais hors normes, ainsi que la pertinence de l'approche « intervention » dans le contexte des modèles linéaires dynamiques.

Les changements de niveaux sont le second type d'intervention mis en œuvre ; leurs effets sur le modèle sont tout aussi importants. En laissant la possibilité au niveau moyen d'effectuer un saut structurel lorsqu'il est justifié, les écarts des observations au modèle sont réduites, et avec eux la variance d'observation ce qui affecte le « ratio signal sur bruit » et ainsi l'adaptabilité et finalement l'ajustement du modèle aux données ce qui se traduit par une augmentation de la log-vraisemblance. Au-delà de cet aspect mécanique de l'intervention, le changement de niveau constitue surtout un bouleversement dans l'histoire portée par la série temporelle, qui nécessite d'être étayé, sans doute plus que les données exceptionnelles qui, au regard de l'histoire de la série, présentent un caractère anecdotique. De ce point de vue « récit », en premier lieu il faut noter que le changement de méthode de mesure de chlorophylle *a* en décembre de 2009 (cf. p. 8), n'a pas entraîné de changements de niveaux. En second lieu, l'année 2007 semble présenter un caractère particulier pour les deux paramètres et les deux lieux de surveillance considérés. C'est une année exceptionnelle d'un point de vue météorologique^f : le semestre hivernal et printanier est le plus doux observé depuis 1900 et les 6 mois suivants sont qualifiés de maussades, très pluvieux et peu ensoleillés. Les effets sont signalés dans le bulletin de la surveillance du laboratoire environnement ressource d'Arcachon (Auby et al., 2008) en particulier par des concentrations en chlorophylle *a* élevées et également en nutriments, en particulier en mars et en juin. Pour le lieu de surveillance « Arcachon-Bouée 7 », l'analyse de la série temporelle montre un changement de niveau à l'automne 2007 aussi bien pour la chlorophylle *a* que pour les abondances phytoplanctoniques. Au lieu de surveillance « Teychan bis », les résultats montrent un changement de niveau au printemps pour les abondances. En revanche, la série temporelle de la chlorophylle *a* ne requiert pas de changement de niveau bien que les concentrations soient maximales pendant l'année 2007, aussi bien au printemps qu'à l'automne. Ainsi, il y a synchronisation entre un tableau météorologique annuel singulier et des évolutions de la biomasse et de l'abondance

f. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2007/bilan-de-lannee-2007>

phytoplanctonique nécessitant des changements structurels ou présentant des niveaux inhabituels. Toutefois, en dépit de la proximité géographique, les caractéristiques environnementales et hydrologiques contrastées des deux lieux de surveillance induisent des changements à des périodes différentes. Par ailleurs, les deux paramètres mesurés, proches thématiquement, présentent des changements au même semestre à « Arcachon-Bouée 7 », mais ce n'est pas le cas à « Teychan bis ».

Les changements structurels observés en 2007 identifient des périodes pendant lesquelles les niveaux moyens sont plus élevés. Ils mettent en œuvre deux interventions : une première intervention au début de la période se traduisant par une augmentation de niveau et une seconde à la fin de la période se traduisant par une diminution. Un changement structurel peut ne concerner qu'une augmentation ou une diminution. A ce titre, l'analyse de la série temporelle des abondances au lieu de surveillance « Teychan bis » est la plus riche des quatre séries. En effet, elle comporte une intervention suggérée par le modèle au 3 février 2009, la mesure précédente étant au 20 janvier. Or, entre le 23 et le 25 janvier 2009 la tempête exceptionnelle Klaus a touché le sud-ouest de la France^g :

Dès l'approche du centre dépressionnaire, les vents se sont renforcés à proximité des côtes charentaises et aquitaines, les rafales dépassant localement 170 km/h. [...]La durée pendant laquelle les vents exceptionnels ont persisté est, elle aussi, tout à fait remarquable [...] L'intensité exceptionnelle de cette tempête est comparable à celle de la seconde tempête de décembre 1999 qui avait touché le sud de la France les 27 et 28 du mois.

Les abondances phytoplanctoniques sont de l'ordre de 350 000 cell./L au 20 janvier et passent à 8 500 cell./L au 3 février. Ultérieurement les concentrations restent faibles, plus faibles que celles habituellement observées, il faut attendre le 30 juillet pour observer une concentration supérieure à celle du 20 janvier, de l'ordre de 400 000 cell./L. Un niveau moyen comparable à celui de début janvier 2009 n'est pas atteint avant l'automne 2010 (cf. p. 42), soit une période de 21 mois illustrant la résilience de l'écosystème d'Arcachon. Pour mémoire, ce changement de niveau du début de l'année 2009 ne peut pas être attribué au changement de niveau qualitatif de profondeur de prélèvement signalé page 8, qui a lieu début 2008 et qui ne donne pas lieu à intervention. Un second changement structurel est suggéré par le modèle au 6 septembre 2011 : les concentrations passent en 15 jours de l'ordre de 680 000 cell./L (i.e. 23 août) à 31 000 cell./L. Nous n'avons pas trouvé d'événements synchronisés avec cette baisse brutale de concentration. Par ailleurs, le changement de la personne responsable des comptages phytoplanctoniques n'est pas en cause dans la mesure où il intervient au premier comptage de l'année 2012 (cf. p. 7). Les concentrations se maintiennent jusqu'au 17 janvier 2012 à moins de 100 000 cell./L, à l'exception du 20 octobre 2011 et d'un point de vue météorologique, le mois de septembre est identifié comme particulièrement chaud et remarquablement sec^h. Finalement, le niveau moyen reste faible toute l'année 2012 et une partie de l'année 2013, puis l'abondance croît à nouveau (cf. p. 42).

L'identification d'événements synchronisés avec les changements structurels suggérés n'a pas la prétention d'établir des relations de causalité. La congruence soulignée permet au mieux de désigner des pistes d'explications à explorer. En effet, un événement (e.g. semestre clément, tempête) est la réunion d'un ensemble de conditions particulières qui contribuent à part inégales à un choc structurel. De fait, des événements en apparence similaires n'induisent pas nécessairement les mêmes effets. Ainsi, la tempête Xynthia a frappé les côtes atlantiques les 27 et 28 février 2010 : les séries étudiées ici ne semblent pas avoir été affectées. L'événement est resté dans les esprits par les drames qui lui sont associés. Mais d'une part, cette tempête n'est pas aussi exceptionnelle au plan

g. <http://www.meteofrance.fr/contenus-archives/bilan-2009/bilan-de-lannee-2009#>

h. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/autres-annees/bilan-de-lannee-2011#>

météorologique que Klausⁱ, et d'autre part la zone impactée concerne plus la Vendée et la Charente-Maritime que la côte atlantique de la Gironde. Enfin, les submersions exceptionnelles associées à Xynthia sont le fait de la phase avec une marée haute à fort coefficient. Ainsi, dans un événement de ce type, les éléments tels que la force et la direction du vent, la durée de la tempête, les précipitations, le coefficient et la phase avec la marée sont autant de paramètres susceptibles d'affecter la réponse de l'écosystème.

L'ambition de la modélisation structurelle des séries temporelles, incluant l'identification et le traitement approprié des données exceptionnelles et des changements de niveaux soutenus préférentiellement par des information exogènes, est d'obtenir les modèles les plus vraisemblables possible non seulement d'un point de vue statistique mais également d'un point de vue biologique et écologique en les soumettant à l'approbation des experts ayant la connaissance du milieu. C'est cette démarche qui amène à considérer que les quantités estimées dans le cadre des modèles témoignent sans doute d'une réalité des écosystèmes. Dans cette étude, c'est la pertinence des estimations des variances d'observation des séries qui est de premier intérêt. En avant-propos, il faut garder en tête que le plan d'échantillonnage et en particulier la fréquence d'échantillonnage, l'étendue de la période temporelle considérée et les événements intervenant pendant cette période, en nombre et intensité, ont une influence sur les valeurs estimées des paramètres des modèles, notamment des variances. Ce point rappelé, dans les résultats présentés les variances d'observation estimées se sont révélées non-significativement différentes entre les lieux pour chaque paramètre. Ainsi, en première approche, bien que les conditions contrastées aient engendré des réponses différenciées des écosystèmes, les ordres de grandeur de variabilité peuvent être considérés comme similaires. Pour ce qui est de la chlorophylle *a*, au regard de l'approche choisie d'expression de la variabilité, la part structurelle, *i.e.* la tendance et la saisonnalité, est responsable de 17 à 22 % des étendues des intervalles de confiance à 95 % des observations, alors que la part liée à la variance d'observation, *i.e.* par convention ici, toute la variabilité depuis le prélèvement jusqu'à la saisie, y compris la variabilité environnementale, est comprise entre 78 et 83 %. La plus grande partie de la variabilité des observations est due à la variabilité spatiale, temporelle et à l'ensemble du processus d'acquisition de la donnée. Exprimée en précision, calculée en pourcentage de la médiane de la distribution estimée par le modèle, la variance d'observation porte la borne inférieure de l'intervalle de confiance à 95 % des observations à -39 et -43 % de la médiane et la borne supérieure à +110 et +130 % de la médiane. En considérant des ordres de grandeur de ce valeurs de -50 % et +100 %, et une vraie valeur de la médiane de la distribution de 2 µg/L, l'intervalle de confiance à 95 % des observations est [1 ; 4] µg/L, cet intervalle étant sous-estimé, ne prenant pas en compte la part de variabilité associée à la somme de la tendance et de la saisonnalité. Ces variances d'observation pour la chlorophylle *a*, appréhendées à travers ces expressions de variabilité, sont compatibles avec l'expérience des experts. Pour ce qui est de l'abondance phytoplanctonique exprimée en log₁₀ cell./L, au lieu de surveillance « Arcachon-Bouée 7 », la part structurelle est de 21 % et la part environnementale et liée au processus d'observation est de 79 %. En revanche, à « Teychan bis », la part structurelle est de 31 % impliquant une part liée à la variance d'observation de 69 %. Il faut souligner que cette répartition ne diffère que de 1 % par rapport à celle du premier modèle, c'est-à-dire sans aucune intervention, et qu'elle n'a pas été modifiée dans les différentes étapes du processus d'expertise (résultats non présentés). Ainsi, le signal structurel est plus fort à « Teychan bis » qu'à « Arcachon-Bouée 7 ». Comme les événements exogènes relevés précédemment en synchronisation avec des changements de niveaux se sont exprimés de manière différentes sur les deux lieux (*cf.* 2007 et 2009), il ne s'agit pas d'une sensibilité plus importante de « Teychan bis » mais d'une caractéristique propre au lieu :

i. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2010/la-tempete-xynthia-des-2728-fevrier-2010>

globalement, compte tenu d'une période étudiée identique, ce qui se passe sur les deux lieux est différent et, spécifiquement pour l'abondance phytoplanctonique, la force du signal est plus importante à « Teychan bis ». Finalement, l'expression des variances d'observation en précision se traduit par $\pm 14.5\%$ à « Arcachon-Bouée 7 » et $\pm 12.7\%$ à « Teychan bis ». Ce sont des valeurs très inférieures à celle constatées pour la chlorophylle *a*, mais ces amplitudes plus réduites sont liées à l'unité, *i.e.* \log_{10} cell./L. Ramenées en cell./L, plus parlantes pour les personnels en charge des comptages phytoplanctoniques, les résultats sont de -53% et $+648\%$ pour « Arcachon-Bouée 7 » et -39% $+696\%$ pour « Teychan bis ». Ces amplitudes sont compatibles avec l'expérience des experts. Outre les variabilités temporelles et spatiales que l'on sait être potentiellement très importantes, le processus de comptage contribue sans doute également de manière significative à cette variabilité. En particulier, lors de comptages de transects de cuve effectués pour des espèces en fortes abondances, les conversions en cell./L impliquent des multiplicateurs importants : ainsi de faibles variations lors du comptage peuvent induire des différences importantes de concentrations.

Plusieurs démarches méthodologiques sont envisageables afin d'améliorer et de prolonger les résultats présentés ici. Les transformations appliquées aux variables pour stabiliser les variances et ainsi satisfaire aux hypothèses du modèle, impliquent une relation quadratique entre moyenne et variance. C'est une hypothèse forte qui pourrait être assouplie en utilisant un modèle à variance variable dans le temps. Une autre possibilité permettant de contourner la transformation de la variable dépendante serait la mise en œuvre de modèles non-gaussiens. Cette approche reste compatible avec la possibilité d'une variance variable dans le temps. Ce dernier raffinement implique l'abandon de l'estimation par maximum de vraisemblance des paramètres du modèle au profit de méthodes procédant par simulation (*e.g.* *Gibbs sampler*). Ces techniques ouvrent alors la porte à des modèles plus complexes. En particulier, l'approche actuelle procède séquentiellement par l'ajustement de modèles successifs, les résultats de l'un étant utilisés, le cas échéant à l'issue d'un choix pouvant contenir une part subjective, pour spécifier le suivant. Petris et al. (2009) décrivent un modèle *for outliers and structural breaks* qui permet en un seul processus d'estimation, d'identifier et de traiter les valeurs exceptionnelles et changements structurels pour toutes les composantes du modèle y compris la saisonnalité. Fúquene et al. (2015) ont publié les résultats d'une version robuste de ce type de modèle.

Le traitement parallèle de plusieurs lieux de surveillance et de plusieurs paramètres permet de confirmer le caractère singulier d'une période temporelle. Par exemple, au lieu de surveillance « Arcachon-Bouée 7 », abondance et biomasse présentent des changements de niveaux synchronisés mais c'est l'existence d'un changement sur la série de chlorophylle *a* qui a en partie motivé la suggestion d'un changement sur la série des abondances. Plus globalement, l'étude simultanée de différents lieux permet d'enrichir les expertises de l'ensemble des lieux. Au-delà du traitement en parallèle de plusieurs lieux de surveillance, il est envisageable de traiter conjointement les lieux pour tirer bénéfice des corrélations qui peuvent exister dans les évolutions des paramètres étudiées. Le cadre méthodologique de cette approche est fourni par les modèles SUTSE, *i.e.* *Seemingly Unrelated Time Series Equations*. Avec ces modèles, ce ne sont pas les séries brutes, très bruitées, pour lesquelles des corrélations sont introduites mais ce sont pour les paramètres des modèles, au premier chef desquels, le niveau moyen et la saisonnalité. Par ailleurs, en ajoutant une composante hiérarchique à la spécification de tels modèles, sous la forme d'une variable aléatoire unique et commune à tous les lieux, on ouvre la possibilité de distinguer dans la variance d'observation une part commune à tous les sites et une part spécifique à chacun d'eux. Lorsqu'un laboratoire unique est en charge des mesures pour les différents sites considérés, alors la part commune à tous les sites peut être interprétée comme la variance induite par le laboratoire, *i.e.* uniquement les opérations de prélèvement et de

laboratoire, alors que les parts spécifiques à chaque site témoignent de la variabilité environnementale propre à chaque lieu de surveillance.

Le traitement parallèle de plusieurs variables d'un même lieu permet d'enrichir le traitement de chacune des séries. Ces variables peuvent être thématiquement proches, comme c'est le cas avec l'abondance et la biomasse, mais également avec d'autres variables dérivées des comptages phytoplanctoniques comme par exemple les abondances des ochrophyta, des myzozoa, le ratio des abondances de ces deux phylums, la richesse spécifique, la richesse des ochrophyta d'une part et celle des myzozoa d'autre part. Les variables peuvent également être plus causales comme par exemple les concentrations en nutriments, leurs ratios, température, salinité, turbidité et matières en suspensions, force et direction du vent, valeurs des coefficients de marées. La prise en compte conjointe de ces variables explicatives peut se faire dans le cadre d'une régression dynamique. Le cadre posé par les modèles SUTSE est également pertinent, surtout compte tenu des variabilités des données brutes.

5. CONCLUSION

L'application de modèles linaires dynamiques aux séries temporelles de chlorophylle *a* et d'abondance phytoplanctonique sur les lieux de surveillance « Arcachon-Bouée 7 » et « Teychan bis » a permis d'estimer leurs variances d'observation, *i.e.* la part de variabilité depuis le prélèvement jusqu'à la saisie des données dans la base Quadrige, intégrant notamment la variabilité environnementale. Ces variances se sont révélées non significativement différentes entre les lieux pour les deux paramètres considérés. Pour la chlorophylle *a*, la variabilité d'observation est responsable de l'ordre de 80 % de la variabilité totale. Pour l'abondance totale, elle est également de 80 % à « Arcachon-Bouée 7 » mais de l'ordre de 70 % à « Teychan bis ». Ainsi la part de « bruit » dépend du lieu et du paramètre considéré. Exprimée en pourcentage de la médiane de la chlorophylle *a*, la variance d'observation place les bornes de l'intervalle de confiance à 95 % des observations à des valeurs de l'ordre de -40 % et +120 % de la médiane, l'intervalle étant sous-estimé car ne prenant pas en compte la variabilité structurelle. Pour l'abondance phytoplanctonique en \log_{10} cell./L, les ordres de grandeur correspondant en pourcentage de la moyenne sont de ± 13.5 %. Pour les deux paramètres, ces valeurs sont compatibles avec l'expérience des experts.

Ainsi, l'approche mise en œuvre s'est avérée riche d'enseignement en matière d'incertitude de mesure mais également en ce qui concerne les évolutions des séries temporelles étudiées, en particulier par la mise en relation de changements de niveau dans les séries avec des événements météorologiques majeurs. Toutefois, l'application méthodologique réalisée ici apparaît n'être que la partie émergente d'un ensemble plus vaste ouvrant des perspectives fécondes à tout point de vue.

6. BIBLIOGRAPHIE

- Alverson, A.J., Manoylov, K.M., Stevenson, R.J., 2003. Laboratory sources of error for algal community attributes during sample preparation and counting. *J. Appl. Phycol.* 15, 357-369.
- Aminot, A., Kérouel, R., 2004. Hydrologie des écosystèmes marins : paramètres et analyses, Méthodes d'analyse en milieu marin. Ifremer.
- Auby, I., Cantin, C., Neaud-Masson, N., 2008. Qualité du Milieu Marin Littoral Bulletin de la surveillance. Départements : Gironde, Landes, Pyrénées Atlantiques.

- Brown, C.D., Canfield, D.E., Bachmann, R.W., Hoyer, M.V., 1998. Seasonal Patterns of Chlorophyll, Nutrient Concentrations and Secchi Disk Transparency in Florida Lakes. *J. Lake Reserv. Manag.* 14, 60-76.
- Cloern, J.E., Powell, T.M., Huzzey, L.M., 1989. Spatial and temporal variability in South San Francisco Bay (USA). II. Temporal changes in salinity, suspended sediments, and phytoplankton biomass and productivity over tidal time scales. *Estuar. Coast. Shelf Sci.* 28, 599-613.
- Dickey, T.D., 2003. Emerging ocean observations for interdisciplinary data assimilation systems. *J. Mar. Syst.* 40-41, 5-48. doi:10.1016/S0924-7963(03)00011-3
- Fúquene, J., Álvarez, M., Raúl Pericchi, L., 2015. A robust Bayesian dynamic linear model for Latin-American economic time series: “the Mexico and Puerto Rico cases.” *Lat. Am. Econ. Rev.* 24, 1-17. doi:10.1007/s40503-015-0020-z
- Gagnon, M., Lacroix, G., 1981. The effects of tidal advection and mixing on the statistical dispersion of zooplankton. *J. Exp. Mar. Biol. Ecol.* 56, 9-22.
- Harvey, A., Koopman, S.J., Penzer, J., 1998. Messy time series: a unified approach. *Adv. Econom.* 13, 103-143.
- Harwood, J., Stokes, K., 2003. Coping with uncertainty in ecological advice: lessons from fisheries. *TRENDS Ecol. Evol.* 18, 617-622.
- Hefley, T.J., Tyre, A.J., Blankenship, E.E., 2013. Fitting population growth models in the presence of measurement and detection error. *Ecol. Model.* 263, 244-250.
- Hernández-Fariñas, T., 2015. Analyse et modélisation des évolutions à long terme de la biodiversité phytoplanctonique dans les zones côtières sous l’effet des pressions environnementales et anthropiques (Thèse de doctorat). Nantes.
- Neaud-Masson, N., 2015. Observation et dénombrement du phytoplancton marin par microscopie optique photonique - Spécifications techniques et méthodologiques appliquées au REPHY. Document de méthode.
- Neveux, J., Panouse, M., 1987. Spectrofluorometric determination of chlorophylls and pheophytins. *Arch. Hydrobiol.* 109, 567-581.
- Petris, G., Petrone, S., Campagnoli, P., 2009. *Dynamic linear models with R, Use R!* Springer.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.
- Stoffer, D.S., Tolo, C.M.C., 1992. A note on the Ljung-Box-Pierce portmanteau statistic with missing data. *Stat. Probab. Lett.* 13, 391-396.
- Utermöhl, H., 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. *Mitt Int Ver Theor Angew Limnol* 9, 1-38.
- Venrick, E., 1978a. Sampling strategies, in: *Phytoplankton Manual, Monographs on Oceanographic Methodology.* A. Sournia, pp. 7-16.
- Venrick, E., 1978b. The implications of subsampling, in: *Phytoplankton Manual, Monographs on Oceanographic Methodology.* A. Sournia, pp. 75-87.
- Venrick, E., 1978c. How many cells to count?, in: *Phytoplankton Manual, Monographs on Oceanographic Methodology.* A. Sournia, pp. 167-180.
- West, M., Harrison, J., 1997. *Bayesian forecastind and dynamic models.* Springer-Verlag.

Zarauz, L., Irigoien, X., 2008. Effects of Lugol's fixation on the size structure of natural nano-microplankton samples, analyzed by means of an automatic counting method. *J. Plankton Res.* 30, 1297-1303.

7. ANNEXE 1 : TABLE DE CORRESPONDANCE TAXONS-PHYLUM

La table ci-dessous établit la correspondance entre les taxons saisis dans la base de données Quadrige et les phylums de ces taxons. Parmi ces derniers, l'abondance totale des cellules phytoplanctonique a été approchée par la somme des abondance Ochrophyta (*i.e.* diatomées) et Myzozoa (*i.e.* dinoflagellés).

Taxons	Phylum	Taxons	Phylum	Taxons	Phylum
Acanthoica	Haptophyta	Bacillaria	Ochrophyta	Chaetoceros didymus var. protuberans	Ochrophyta
Achnantheaceae	Ochrophyta	Bacillaria paxillifer	Ochrophyta		
Achnanthes	Ochrophyta	Bacillaria paxillifera	Ochrophyta	Chaetoceros fragilis	Ochrophyta
Achnanthes brevipes	Ochrophyta	Bacillariaceae	Ochrophyta	Chaetoceros fragilis + wighamii	Ochrophyta
Actinastrum	Chlorophyta	Bacillariophyceae	Ochrophyta		
Actinocyclus	Ochrophyta	Bacteriastrium	Ochrophyta	Chaetoceros lorenzianus	Ochrophyta
Actinoptychus	Ochrophyta	Bellerochea	Ochrophyta	Chaetoceros peruvianus	Ochrophyta
Actinoptychus senarius	Ochrophyta	Biddulphia	Ochrophyta	Chaetoceros protuberans	Ochrophyta
Actinoptychus senarius + campanulifer	Ochrophyta	Biddulphia alternans	Ochrophyta	Chaetoceros pseudocurvisetus	Ochrophyta
Adenoides	Myzozoa	Biddulphia granulata	Ochrophyta	Chaetoceros rostratus	Ochrophyta
Akashiwo	Myzozoa	Biddulphia rhombus	Ochrophyta	Chaetoceros socialis	Ochrophyta
Akashiwo sanguinea	Myzozoa	Biddulphiaceae	Ochrophyta	Chaetoceros socialis + socialis f. radians	Ochrophyta
Alexandrium	Myzozoa	Bleakeleya	Ochrophyta		
Alexandrium affine	Myzozoa	Blepharocysta	Myzozoa	Chaetoceros socialis f. radians	Ochrophyta
Alexandrium andersonii	Myzozoa	Brockmanniella	Ochrophyta	Chaetoceros socialis f. radians	Ochrophyta
Alexandrium catenella	Myzozoa	Brockmanniella brockmannii	Ochrophyta	Chaetoceros subtilis	Ochrophyta
Alexandrium hiranoi	Myzozoa	Campylsira	Ochrophyta	Chaetoceros wighamii	Ochrophyta
Alexandrium insuetum	Myzozoa	Campylsira cymbelliformis	Ochrophyta	Chaetocerotaceae	Ochrophyta
Alexandrium leei	Myzozoa	Centrales	Ochrophyta	Chattonella	Ochrophyta
Alexandrium margalefii	Myzozoa	Centriques	Ochrophyta	Chattonellales	Ochrophyta
Alexandrium minutum	Myzozoa	Cerataulina	Ochrophyta	Chlorodendrales	Chlorophyta
Alexandrium ostenfeldii	Myzozoa	Cerataulina pelagica	Ochrophyta	Chlorophyceae	Chlorophyta
Alexandrium pseudogonyaulax	Myzozoa	Ceratium	Myzozoa	Choanofila	Choanozoa
Alexandrium tamarense	Myzozoa	Ceratium + Neoceratium	Myzozoa	Chrysochromulina	Haptophyta
Alexandrium tamarense + catenella + tamutum	Myzozoa	Ceratium furca	Myzozoa	Chrysochromulina hirta + ericina + spinifera	
Alexandrium tamutum	Myzozoa	Ceratium fusus	Myzozoa	Chrysophyceae	Ochrophyta
Amphidiniopsis	Myzozoa	Ceratium horridum	Myzozoa	Ciliophora	Ciliophora
Amphidinium	Myzozoa	Ceratium lineatum	Myzozoa	Climaconeis	Ochrophyta
Amphidinium carterae	Myzozoa	Ceratium lineatum + minutum	Myzozoa	Climacosphenia	Ochrophyta
Amphidinium carterae + operculatum	Myzozoa	Ceratium macroceros	Myzozoa	Closterium	Streptophyta
Amphidinium crassum	Myzozoa	Ceratium pentagonum	Myzozoa	Coccolithaceae	Haptophyta
Amphidoma	Myzozoa	Ceratium ranipes	Myzozoa	Coccolithus	Haptophyta
Amphidoma caudata	Myzozoa	Ceratium teres	Myzozoa	Cocconeis	Ochrophyta
Amphipora	Ochrophyta	Ceratium tripos	Myzozoa	Coccosphaerales	Haptophyta
Amphora	Ochrophyta	Ceratium tripos + Ceratium à cornes recourbées	Myzozoa	Cochlodinium	Myzozoa
Amylax	Myzozoa	Ceratocorys	Myzozoa	Cochlodinium polykrikoides	Myzozoa
Amylax triacantha	Myzozoa	Ceratoneis closterium	Ochrophyta	Coolia	Myzozoa
Amylax triacantha + buxus	Myzozoa	Chaetoceros	Ochrophyta	Corethron	Ochrophyta
Anabaena	Cyanobacteria	Chaetoceros affinis	Ochrophyta	Corethron criophilum	Ochrophyta
		Chaetoceros armatus	Ochrophyta	Corethron pennatum	Ochrophyta
		Chaetoceros compressus	Ochrophyta	Corythodinium	Myzozoa
		Chaetoceros curvisetus	Ochrophyta	Coscinodiscaceae	Ochrophyta
Ankistrodesmus	Chlorophyta	Chaetoceros curvisetus + debilis + pseudocurvisetus	Ochrophyta	Coscinodiscophyceae	Ochrophyta
Apedinella radians	Ochrophyta	Chaetoceros danicus	Ochrophyta	Coscinodiscophycidae	Ochrophyta
Archaeoperidinium minutum	Myzozoa	Chaetoceros debilis	Ochrophyta	Coscinodiscus	Ochrophyta
Asterionella	Ochrophyta	Chaetoceros decipiens	Ochrophyta	Coscinodiscus + Stellarima	Ochrophyta
Asterionella + Asterionellopsis + Asteroplanus	Ochrophyta	Chaetoceros decipiens + lorenzianus	Ochrophyta	Coscinodiscus asteromphalus + oculus-iridis + perforatus	Ochrophyta
Asterionella formosa	Ochrophyta	Chaetoceros densus	Ochrophyta	Coscinodiscus granii	Ochrophyta
Asterionellopsis	Ochrophyta	Chaetoceros densus + castracanei	Ochrophyta	Coscinodiscus oculus-iridis	Ochrophyta
Asterionellopsis glacialis	Ochrophyta	Chaetoceros densus + impressus	Ochrophyta	Coscinodiscus oculus-iridis	Ochrophyta
Asterolampraceae	Ochrophyta	Chaetoceros densus + impressus	Ochrophyta	Coscinodiscus radiatus + marginatus	Ochrophyta
Asteromphalus	Ochrophyta	Chaetoceros diadema	Ochrophyta	Coscinodiscus walesii	Ochrophyta
Asteroplanus karianus	Ochrophyta	Chaetoceros didymus	Ochrophyta	Cosmarium	Streptophyta
Attheya	Ochrophyta	Chaetoceros didymus	Ochrophyta	Crucigenia	Chlorophyta
Attheya armata	Ochrophyta	Chaetoceros didymus + didymus var. protuberans	Ochrophyta	Cryptomonadales	Cryptophyta
Attheya armatus	Ochrophyta	Chaetoceros didymus + protuberans	Ochrophyta	Cryptophyceae	Cryptophyta
Aulacodiscus	Ochrophyta			Cyanobacteria	Cyanobacteria
Aulacoseira	Ochrophyta				ia

Taxons	Phylum	Taxons	Phylum	Taxons	Phylum
Cyanophyceae	Cyanobacteria	Euglenida	Euglenozoa	Katodinium glaucum	Myzozoa
		Euglenoidea	Euglenozoa	Kryptoperidinium foliaceum	Myzozoa
Cyclotella	Ochrophyta	Eutreptia	Euglenozoa	Lauderia	Ochrophyta
Cylindrotheca	Ochrophyta	Eutreptiaceae	Euglenozoa	Lauderia + Schroederella	Ochrophyta
Cylindrotheca closterium	Ochrophyta	Eutreptiales	Euglenozoa	Lauderia annulata	Ochrophyta
Cylindrotheca gracilis	Ochrophyta	Eutreptiella	Euglenozoa	Lepidodinium	Myzozoa
Cymatosiraceae	Ochrophyta	Eutreptiida	Euglenozoa	Lepidodinium chlorophorum	Myzozoa
Cymbella	Ochrophyta	Fibrocapsa	Ochrophyta	Leptocylindraceae	Ochrophyta
Cymbellaceae	Ochrophyta	Fibrocapsa japonica	Ochrophyta	Leptocylindrus	Ochrophyta
Dactyliosolen	Ochrophyta	Fragilaria	Ochrophyta	Leptocylindrus danicus	Ochrophyta
Dactyliosolen fragilissimus	Ochrophyta	Fragilariaceae	Ochrophyta	Leptocylindrus danicus +	Ochrophyta
Delphineis	Ochrophyta	Fragilariopsis	Ochrophyta	curvatus	
Denticula	Ochrophyta	Fragilidium	Myzozoa	Leptocylindrus danicus +	Ochrophyta
Detonula	Ochrophyta	Gomphonema	Ochrophyta	curvatus	
Detonula pumila	Ochrophyta	Gomphonema + Gomphoneis	Ochrophyta	Leptocylindrus	Ochrophyta
Diatoma	Ochrophyta	Goniodoma	Myzozoa	mediterraneus	
Dictyocha	Ochrophyta	Goniodoma polyedricum	Myzozoa	Leptocylindrus minimus	Ochrophyta
Dictyocha fibula	Ochrophyta	Goniodomataceae	Myzozoa	Licmophora	Ochrophyta
Dictyocha speculum	Ochrophyta	Gonyaulacaceae	Myzozoa	Lingulodinium	Myzozoa
Dictyochaceae	Ochrophyta	Gonyaulax	Myzozoa	Lingulodinium polyedrum	Myzozoa
Dictyochales	Ochrophyta	Gonyaulax grindleyi	Myzozoa	Lithodesmiaceae	Ochrophyta
Dictyochophyceae	Ochrophyta	Gonyaulax spinifera	Myzozoa	Lithodesmium	Ochrophyta
Dinobryon	Ochrophyta	Gonyaulax verior	Myzozoa	Lithodesmium undulatum	Ochrophyta
Dinoflagellata	Myzozoa	Grammatophora	Ochrophyta	Lyrella	Ochrophyta
Dinophyceae	Myzozoa	Grammatophora oceanica	Ochrophyta	Mediopyxis helysia	Ochrophyta
Dinophysiaceae	Myzozoa	Grammatophora serpentina	Ochrophyta	Melosira	Ochrophyta
Dinophysiales	Myzozoa	Guinardia	Ochrophyta	Melosira moniliformis	Ochrophyta
Dinophysida	Myzozoa	Guinardia delicatula	Ochrophyta	Melosira nummuloides	Ochrophyta
Dinophysis	Myzozoa	Guinardia flaccida	Ochrophyta	Melosiraceae	Ochrophyta
Dinophysis acuminata	Myzozoa	Guinardia striata	Ochrophyta	Merismopedia	Cyanobacteria
Dinophysis acuta	Myzozoa	Gymnodiniaceae	Myzozoa		
Dinophysis caudata	Myzozoa	Gymnodiniales	Myzozoa	Mesodinium	Ciliophora
Dinophysis diegensis	Myzozoa	Gymnodiniida	Myzozoa	Mesodinium rubrum	Ciliophora
Dinophysis fortii	Myzozoa	Gymnodinium	Myzozoa	Mesoporos	Myzozoa
Dinophysis hastata + odiosa	Myzozoa	Gymnodinium + Gyrodinium	Myzozoa	Metaphalacroma	Myzozoa
Dinophysis norvegica	Myzozoa	Gymnodinium catenatum	Myzozoa	Meuniera	Ochrophyta
Dinophysis rapa	Myzozoa	Gymnodinium impudicum	Myzozoa	Meuniera membranacea	Ochrophyta
Dinophysis sacculus	Myzozoa	Gyrodinium	Myzozoa	Michaelsarsia adriaticus	Haptophyta
Dinophysis skagii	Myzozoa	Gyrodinium spirale	Myzozoa	Micracanthodinium	Myzozoa
Dinophysis tripos	Myzozoa	Gyrosigma	Ochrophyta	Microcystis	Cyanobacteria
Diploneis	Ochrophyta	Gyrosigma fasciola	Ochrophyta		
Diplopsalis	Myzozoa	Halosphaera	Chlorophyta	Minutocellus	Ochrophyta
Diplopsalis+Diplopelta+Diplo	Myzozoa	Hantzschia	Ochrophyta	Navicula	Ochrophyta
psalopsis+Preperidinium+Obl		Haptolina hirta + ericina +	Haptophyta	Navicula + Fallacia + Haslea	Ochrophyta
ea		Chrysochromulina spinifera		+ Lyrella + Petroneis	
Diplopsalis+Diplopelta+Diplo	Myzozoa	Haslea	Ochrophyta	Navicula gregaria	Ochrophyta
psalopsis+Zygabikodinium+O		Haslea ostrearia	Ochrophyta	Navicula gregaria +	Ochrophyta
blea		Haslea wawriake	Ochrophyta	cryptocephala	
Diplopsalopsis	Myzozoa	Helicostomella	Ciliophora	Navicula pelagica	Ochrophyta
Discosphaera	Haptophyta	Helicotheca	Ochrophyta	Naviculaceae	Ochrophyta
Dissodinium	Myzozoa	Helicotheca tamesis	Ochrophyta	Naviculales	Ochrophyta
Dissodinium + Pyrocystis	Myzozoa	Heliopeltaceae	Ochrophyta	Naviculineae	Ochrophyta
Ditylum	Ochrophyta	Hemiaulus	Ochrophyta	Nematodinium	Myzozoa
Ditylum brightwellii	Ochrophyta	Hemiaulus membranaceus	Ochrophyta	Neocalyptrella robusta	Ochrophyta
Donkinia	Ochrophyta	Hemiaulus sinensis	Ochrophyta	Neoceratium candelabrum	Myzozoa
Ebria	Protozoa	Hemidiscaceae	Ochrophyta	Neoceratium declinatum	Myzozoa
		Hermesinum	Protozoa	Neoceratium furca	Myzozoa
			incertae	Neoceratium fusus	Myzozoa
Ebria tripartita	Protozoa		sedis	Neoceratium horridum	Myzozoa
		Heterocapsa	Myzozoa	Neoceratium lineatum	Myzozoa
		Heterocapsa niei	Myzozoa	Neoceratium lineatum +	Myzozoa
Ebriaceae	Protozoa	Heterocapsa rotundata	Myzozoa	minutum	
		Heterocapsa triquetra	Myzozoa	Neoceratium macroceros	Myzozoa
		Heterosigma	Ochrophyta	Neoceratium pentagonum	Myzozoa
Ensiculifera	Myzozoa	Heterosigma akashiwo	Ochrophyta	Neoceratium ranipes	Myzozoa
Entomoneidaceae	Ochrophyta	Karenia	Myzozoa	Neoceratium symmetricum	Myzozoa
Entomoneis	Ochrophyta	Karenia brevis	Myzozoa	Neoceratium teres	Myzozoa
Epithemia	Ochrophyta	Karenia brevis +	Myzozoa	Neoceratium tripos	Myzozoa
Eucampia	Ochrophyta	papilionacea		Nitzschia	Ochrophyta
Eucampia + Climacodium	Ochrophyta	Karenia mikimotoi	Myzozoa	Nitzschia + Hantzschia	Ochrophyta
Eucampia zodiacus	Ochrophyta	Karenia papilionacea	Myzozoa	Nitzschia longissima	Ochrophyta
Euglena	Euglenozoa	Karlodinium	Myzozoa	Nitzschia sp1	Ochrophyta
Euglenaceae	Euglenozoa	Karlodinium veneficum	Myzozoa	Nitzschia sp2	Ochrophyta
Euglenales	Euglenozoa	Katodinium	Myzozoa	Noctiluca	Myzozoa

Taxons	Phylum	Taxons	Phylum	Taxons	Phylum
Noctiluca scintillans	Myzozoa	Prorocentrida	Myzozoa	Pseudo-nitzschia, complexe	Ochrophyta
Noctilucaeae	Myzozoa	Prorocentrum	Myzozoa	delicatissima, groupe des	
Noctilucales	Myzozoa	Prorocentrum balticum +	Myzozoa	fines (calliantha +	
Noctilucida	Myzozoa	cordatum		delicatissima +	
Nostoc	Cyanobacter	Prorocentrum compressum	Myzozoa	pseudodelicatissima +	
	ia	Prorocentrum cordatum	Myzozoa	subcurvata)	
Oblea	Myzozoa	Prorocentrum gracile	Myzozoa	Pseudo-nitzschia, complexe	Ochrophyta
Odontella	Ochrophyta	Prorocentrum lima	Myzozoa	seriata, groupe des effilées	
Odontella aurita	Ochrophyta	Prorocentrum lima +	Myzozoa	(multiseries + pungens)	
Odontella granulata	Ochrophyta	marinum		Pseudo-nitzschia, complexe	Ochrophyta
Odontella mobilensis	Ochrophyta	Prorocentrum mexicanum	Myzozoa	seriata, groupe des larges	
Odontella regia	Ochrophyta	Prorocentrum mexicanum +	Myzozoa	(australis + fraudulenta +	
Odontella rhombus	Ochrophyta	rhathymum		seriata + subpacificica)	
Odontella sinensis	Ochrophyta	Prorocentrum micans	Myzozoa	Pseudo-nitzschia, groupe des	Ochrophyta
Olisthodiscus	Ochrophyta	Prorocentrum micans +	Myzozoa	effilées, complexe seriata	
Ophiaster	Haptophyta	arcuatum + gibbosum		(multiseries + pungens)	
Ornithocercus	Myzozoa	Prorocentrum minimum	Myzozoa	Pseudo-nitzschia, groupe des	Ochrophyta
Ornithocercus magnificus	Myzozoa	Prorocentrum minimum +	Myzozoa	fines, complexe	
Oscillatoria	Cyanobacter	balticum + cordatum		delicatissima (calliantha +	
	ia	Prorocentrum triestinum	Myzozoa	delicatissima +	
Ostreopsis	Myzozoa	Protoceratium	Myzozoa	pseudodelicatissima)	
Ostreopsis ovata	Myzozoa	Protoceratium reticulatum	Myzozoa	Pseudo-nitzschia, groupe des	Ochrophyta
Oxyphysis	Myzozoa	Protoctista	Protoctista	larges asymétriques	
Oxyrrhis	Myzozoa	Protoperidinium	Myzozoa	(australis + seriata +	
Oxyrrhis marina	Myzozoa	Protoperidinium +	Myzozoa	subpacificica)	
Oxytoxaceae	Myzozoa	Peridinium		Pseudo-nitzschia, groupe des	Ochrophyta
Oxytoxum	Myzozoa	Protoperidinium bipes	Myzozoa	larges symétriques	
Oxytoxum + Corythodinium	Myzozoa	Protoperidinium brevipes	Myzozoa	(fraudulenta)	
Pachysphaera + Pterosperma	Chlorophyta	Protoperidinium conicum	Myzozoa	Pseudo-nitzschia, groupe des	Ochrophyta
Palaeophalacroma	Myzozoa	Protoperidinium crassipes	Myzozoa	larges, complexe seriata	
Paralia	Ochrophyta	Protoperidinium depressum	Myzozoa	(australis + fraudulenta +	
Paralia sulcata	Ochrophyta	Protoperidinium diabolium	Myzozoa	seriata + subpacificica)	
Pediastrum	Chlorophyta	Protoperidinium diabolium +	Myzozoa	Pseudo-nitzschia, groupe des	Ochrophyta
Pedinellales	Ochrophyta	longipes		sigmoïdes (multistriata)	
Pennales	Ochrophyta	Protoperidinium diabolus	Myzozoa	Pseudopedinella	Ochrophyta
Pennées	Ochrophyta	Protoperidinium diabolus +	Myzozoa	Pyramimonas	Chlorophyta
Peridiniaceae	Myzozoa	longipes		Pyramimonas longicauda	Chlorophyta
Peridinales	Myzozoa	Protoperidinium divergens	Myzozoa	Pyrocystis	Myzozoa
Peridiniella	Myzozoa	Protoperidinium granii	Myzozoa	Pyrophacaceae	Myzozoa
Peridiniida	Myzozoa	Protoperidinium minutum	Myzozoa	Pyrophacus	Myzozoa
Peridinium	Myzozoa	Protoperidinium oblongum	Myzozoa	Raphidophyceae	Ochrophyta
Peridinium quinquecorne	Myzozoa	Protoperidinium obtusum	Myzozoa	Rhabdolithes	Haptophyta
Phaeocystis	Haptophyta	Protoperidinium pentagonum	Myzozoa	Rhabdonema	Ochrophyta
Phalacroma	Myzozoa	+ latissimum		Rhabdonema adriaticum	Ochrophyta
Phalacroma rapa	Myzozoa	Protoperidinium	Myzozoa	Rhabdosphaera	Haptophyta
Phalacroma rotundata	Myzozoa	punctulatum		Rhaphoneis	Ochrophyta
Phalacroma rotundatum	Myzozoa	Protoperidinium steinii	Myzozoa	Rhaphoneis + Delphineis	Ochrophyta
Phytoflagellés excepté	Chromista	Protoperidinium steinii +	Myzozoa	Rhizosolenia	Ochrophyta
dinoflagellés		pyriforme		Rhizosolenia hebetata	Ochrophyta
Pinnularia	Ochrophyta	Prymnesiaceae	Haptophyta	Rhizosolenia imbricata	Ochrophyta
Plagiogramma	Ochrophyta	Prymnesiales	Haptophyta	Rhizosolenia imbricata +	Ochrophyta
Plagiogrammopsis	Ochrophyta	Prymnesiophyceae	Haptophyta	styliformis	
Plagiogrammopsis	Ochrophyta	Prymnesium	Haptophyta	Rhizosolenia pungens	Ochrophyta
vanheurckii		Pseliodinium	Myzozoa	Rhizosolenia robusta	Ochrophyta
Plagiotropis	Ochrophyta	Pseudo-nitzschia	Ochrophyta	Rhizosolenia setigera	Ochrophyta
Pleurosigma	Ochrophyta	Pseudo-nitzschia americana	Ochrophyta	Rhizosolenia setigera +	Ochrophyta
Pleurosigma + Gyrosigma	Ochrophyta	Pseudo-nitzschia australis	Ochrophyta	pungens	
Podocystis	Ochrophyta	Pseudo-nitzschia	Ochrophyta	Rhizosolenia setigera +	Ochrophyta
Podolampas	Myzozoa	delicatissima		setigera f. pungens	
Podolampas palmipes	Myzozoa	Pseudo-nitzschia fraudulenta	Ochrophyta	Rhizosolenia setigera f.	Ochrophyta
Podosira	Ochrophyta	Pseudo-nitzschia multistriata	Ochrophyta	pungens	
Podosira + Hyalodiscus	Ochrophyta	Pseudo-nitzschia	Ochrophyta	Rhizosolenia styliformis	Ochrophyta
Podosira stelligera	Ochrophyta	pseudodelicatissima		Rhizosoleniaceae	Ochrophyta
Polykrikaceae	Myzozoa	Pseudo-nitzschia pungens	Ochrophyta	Scenedesmus	Chlorophyta
Polykrikos	Myzozoa	Pseudo-nitzschia seriata	Ochrophyta	Scenedesmus quadricauda	Chlorophyta
Polykrikos schwarzii	Myzozoa	Pseudo-nitzschia subpacificica	Ochrophyta	Schroederella	Ochrophyta
Porosira	Ochrophyta	Pseudo-nitzschia, complexe	Ochrophyta	Scrippsiella	Myzozoa
Prasinophyceae	Chlorophyta	americana (americana +		Scrippsiella + Ensiculifera +	Myzozoa
Proboscia	Ochrophyta	brasiliiana)		Pentapharsodinium +	
Proboscia alata	Ochrophyta	Pseudo-nitzschia, complexe	Ochrophyta	Bysmatrum	
Proboscia indica	Ochrophyta	americana (americana)		Selenastrum	Chlorophyta
Pronoctiluca	Myzozoa			Sinophysis	Myzozoa
Prorocentraceae	Myzozoa			Skeletonema	Ochrophyta
Prorocentrales	Myzozoa			Skeletonema costatum	Ochrophyta

Taxons	Phylum
Spatulodinium	Myzozoa
Spatulodinium pseudonociluca	Myzozoa
Staurastrum	Streptophyta
Stauroneis	Ochrophyta
Stephanopyxis	Ochrophyta
Streptotheca	Ochrophyta
Striatella	Ochrophyta
Surirella	Ochrophyta
Surirellaceae	Ochrophyta
Synedra	Ochrophyta
Synedra + Toxarium	Ochrophyta
Syracosphaera	Haptophyta
Tabellaria	Ochrophyta
Tetraselmis + Prasinocladus	Chlorophyta
Thalassionema	Ochrophyta
Thalassionema +	Ochrophyta
Thalassiothrix + Lioloma	
Thalassionema nitzschioides	Ochrophyta
Thalassiosira	Ochrophyta
Thalassiosira + Porosira	Ochrophyta
Thalassiosira angulata	Ochrophyta
Thalassiosira anguste-lineata	Ochrophyta
Thalassiosira antarctica	Ochrophyta
Thalassiosira decipiens	Ochrophyta
Thalassiosira gravida	Ochrophyta
Thalassiosira levanderi	Ochrophyta
Thalassiosira levanderi + minima	Ochrophyta
Thalassiosira nordenskiöldii	Ochrophyta
Thalassiosira punctigera	Ochrophyta
Thalassiosira rotula	Ochrophyta
Thalassiosira rotula + gravida	Ochrophyta
Thalassiosira subtilis	Ochrophyta
Thalassiosiraceae	Ochrophyta
Thalassiothrix	Ochrophyta
Thecadinium	Myzozoa
Tiarina	Ciliophora
Tintinnina	Ciliophora
Torodinium	Myzozoa
Torodinium robustum	Myzozoa
Tous Dinophysis ronds avec épithèque bien visible	Myzozoa
Toxarium undulatum	Ochrophyta
Toxonidea	Ochrophyta
Trachyneis	Ochrophyta
Triceratiaceae	Ochrophyta
Triceratium	Ochrophyta
Triceratium favus	Ochrophyta
Trigonium alternans	Ochrophyta
Tryblionella compressa	Myzozoa
Vulcanodinium rugosum	Myzozoa
Warnowia	Myzozoa
Warnowia + Nematodinium + Nematopsides	Myzozoa
Warnowiaceae	Myzozoa

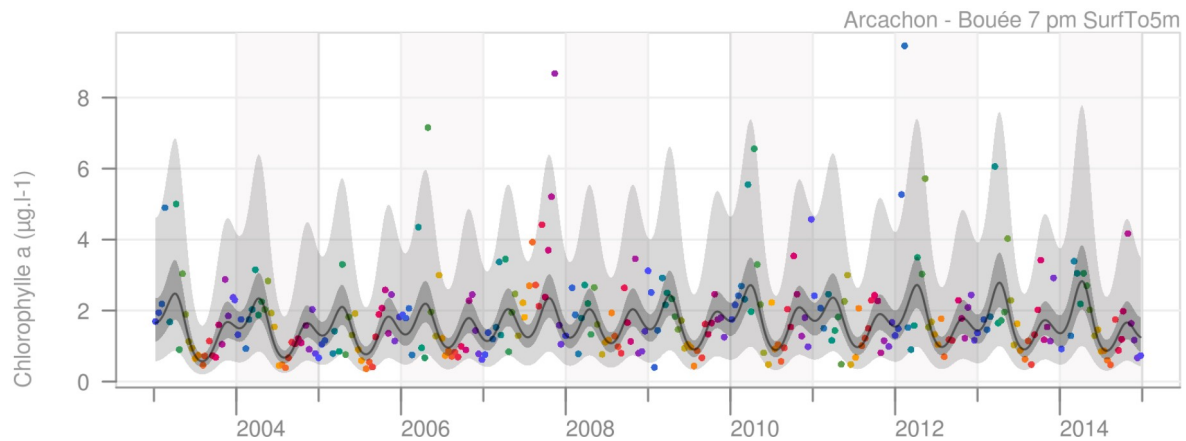
8. ANNEXE 2 : DÉTAILS DES EXPERTISES

8.1 CHLOROPHYLLE A

Pour ce paramètre, la représentation des observations et des résultats du modèle sont dans l'unité native de la mesure : $\mu\text{g/L}$. En revanche, lorsque le niveau moyen du modèle est présenté, comme la modélisation est réalisée sur le logarithme népérien des concentrations, l'unité est alors \log chlorophylle *a* $\mu\text{g/L}$.

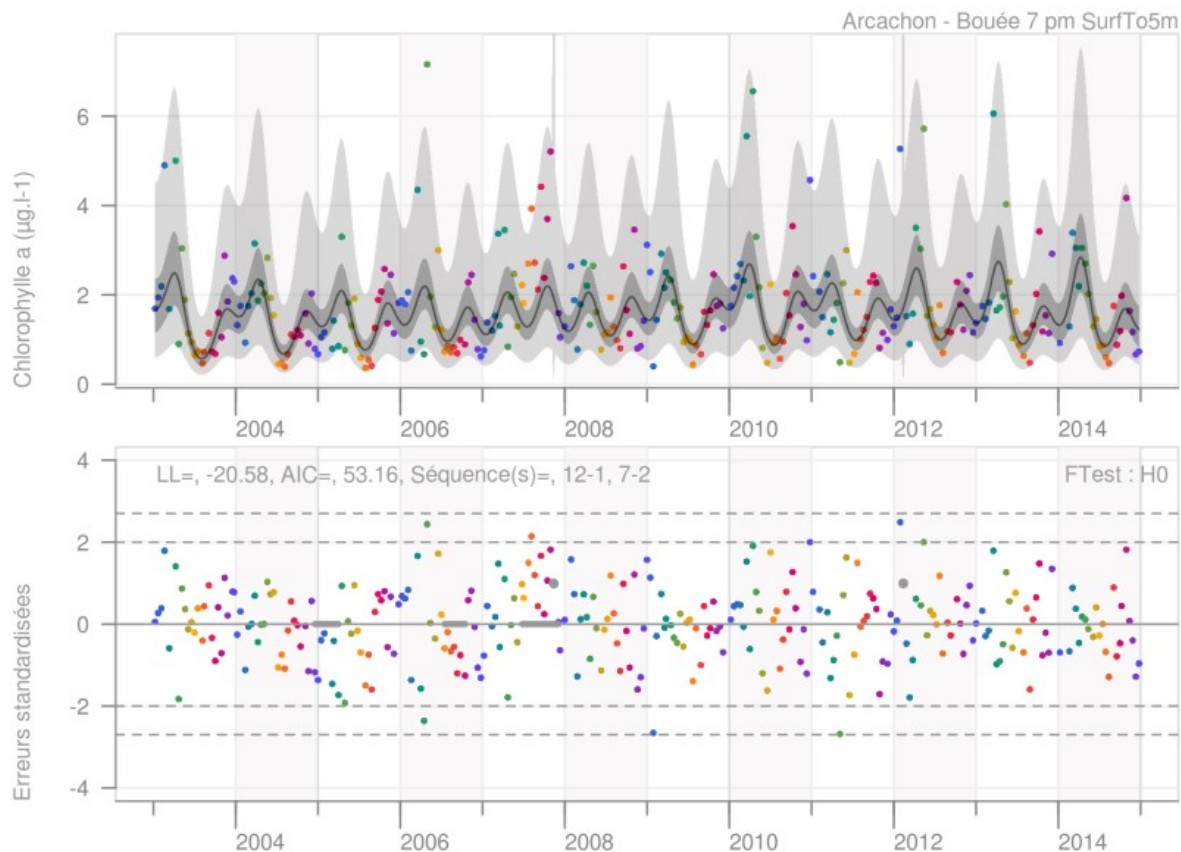
Bouée 7

1. La log-vraisemblance du premier modèle est de -27.6 et la variance d'observation est de 0.247 log de chlorophylle a .



Il suggère deux valeurs exceptionnelles : ce sont les deux concentrations les plus élevées, à plus de 8 $\mu\text{g/L}$ au 19/11/2007 et 16/02/2012.

2. Le second modèle traite ces deux valeurs comme exceptionnelles : sa log-vraisemblance est de -20.58 et le gain très significatif ($p < 0.001$). La variance d'observation diminue à 0.230 log chlorophylle a .



Ce modèle n'identifie pas d'autres valeurs exceptionnelles. En revanche, il identifie trois séquences d'erreurs standardisées de même signe en 2005, 2006 et 2007. Chronologiquement, la première valeur exceptionnelle (19/11/2007) du second modèle est dans une séquence de 12 erreurs standardisées positives consécutives, signe de mauvaise adaptation locale du modèle. La période s'étend du 26 juin 2007

au 19 novembre 2007. Des bilans saisonniers météo-france pour 2007^j on peut retenir :

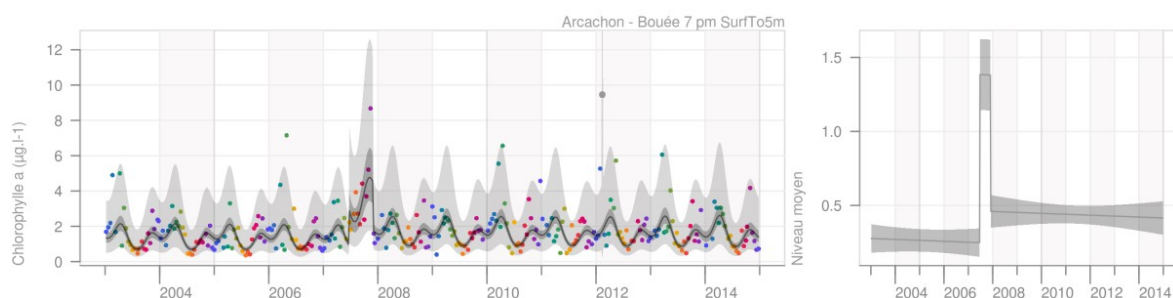
Cet été 2007, globalement très pluvieux et fort peu ensoleillé, s'inscrit dans la lignée des étés maussades de 2002, 1992 et 1977.

Dans le bulletin de la surveillance 2007 (édition 2008) on peut lire :

En 2007, les teneurs en nitrate dans les eaux du Bassin ont été particulièrement élevées par rapport à la normale en mars-avril et en juin-juillet, résultant des forts apports par l'Eyre en mars et en juin [...]

En 2007, des situations assez différentes ont été observées selon la localisation des points dans la Baie. A l'entrée du bassin ("Bouée 7" et "Bouée 11-Bouée 13"), le bloom phytoplanctonique printanier a été observé entre mars et avril. [...] Par la suite, les teneurs en chlorophylle sont restées élevées par rapport aux années précédentes. Comme en 2006, le bloom automnal a été très marqué [...], permettant d'atteindre, en octobre-novembre, des teneurs en chlorophylle très supérieures à la normale.

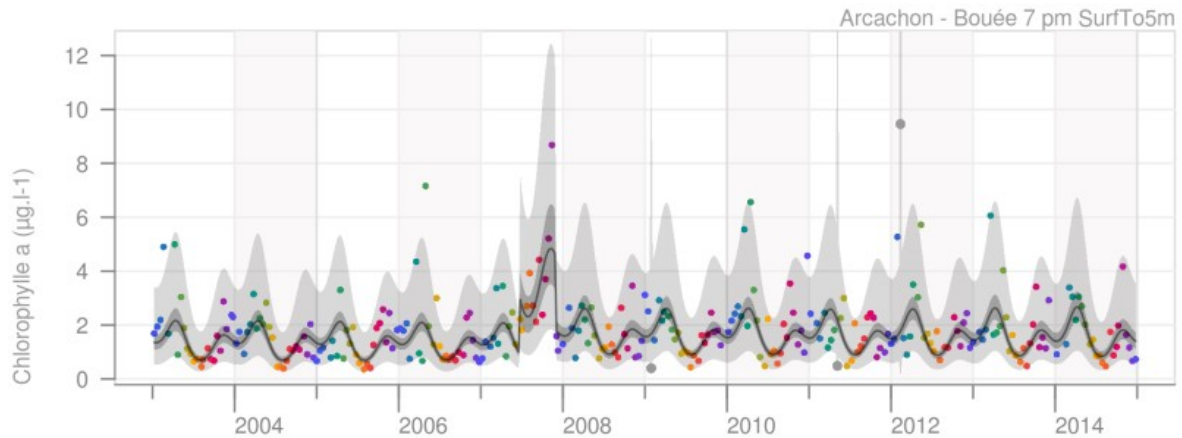
3. Lorsque l'on suggère des changements de niveaux aux bornes de cette période, et ce faisant l'observation du 19/11/2007 n'est plus traitée comme valeur exceptionnelle, l'évolution du niveau moyen est très modifiée. Ce troisième modèle n'étant pas emboîté dans le second modèle, le caractère significatif du gain de log-vraisemblance ne peut pas être testé. Toutefois, il est à noter que la log-vraisemblance du nouveau modèle est de -8.8, que l'AIC de ce dernier est plus petit que pour le second modèle et la variance d'observation de 0.227 log chlorophylle a .



Le modèle suggère encore deux valeurs exceptionnelles : l'une au 3/02/2009, 0.4 µg/L comprise entre une mesure à 2.51 et une autre à 1.44 µg/L. L'autre au 10/05/2011, 0.49 µg/L comprise entre 1.82 et 2.26 µg/L.

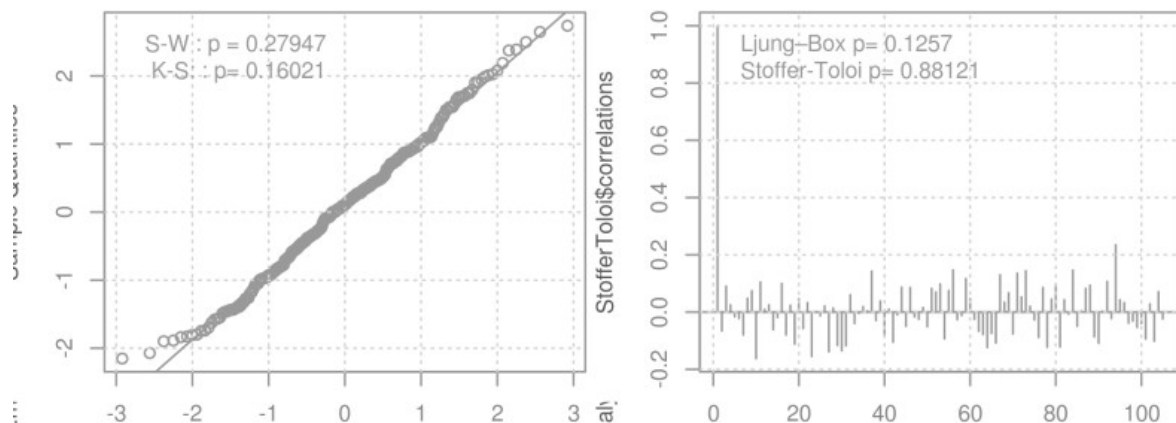
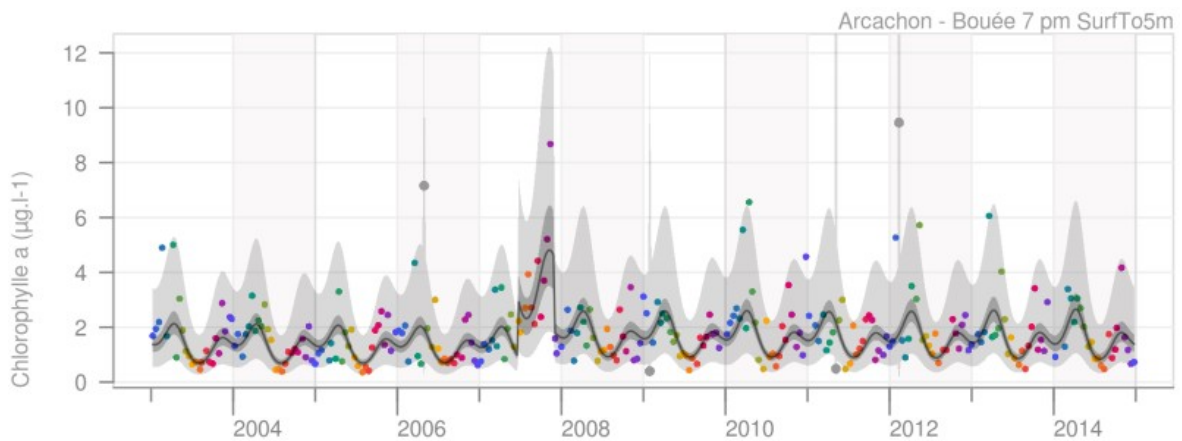
j. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2007/bilan-de-lannee-2007>

4. La log-vraisemblance du nouveau modèle traitant de manière appropriée ces valeurs exceptionnelles est de -2.98 le gain est significatif ($p=0.003$) et la variance d'observation diminue à 0.214 log chlorophylle a .



Ce modèle suggère une dernière valeur exceptionnelle : au 04/05/2006 une valeur de 7.16 $\mu\text{g/L}$ comprise entre 0.67 et 1.96 $\mu\text{g/L}$.

5. Le gain en log-vraisemblance du modèle traitant de manière appropriée cette dernière valeur exceptionnelle est très important ($p < 0.001$, LL = -0.37). La variance d'observation estimée est de 0.208 log chlorophylle a $\mu\text{g/L}$.

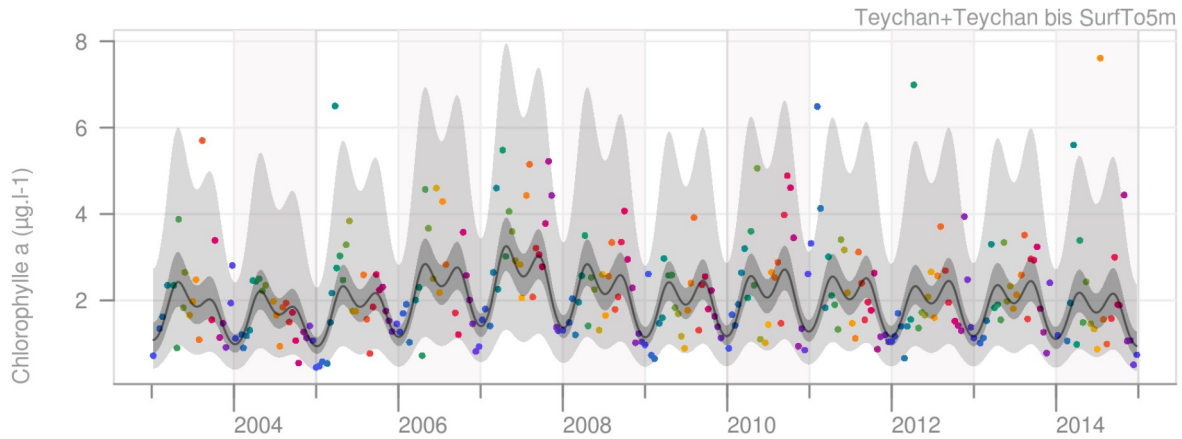


Ce modèle ne suggère pas d'autre valeur exceptionnelle. La normalité des résidus standardisés est acquise aussi bien au regard du Q-Q plot, du test de Kolmogorov-Smirnov, que du Shapiro-Wilks et l'hypothèse de leur indépendance n'est pas rejetée par le test de Stoffer-Toloi. Il reste deux séquences d'erreurs standardisées de même signe. Plusieurs interventions et combinaisons

d'interventions ont été testées pour les dissoudre, mais sans succès probants ni véritable gain en log-vraisemblance. Le dernier modèle est reconnu en l'état comme pertinent par les experts biologique, écologique et terrain.

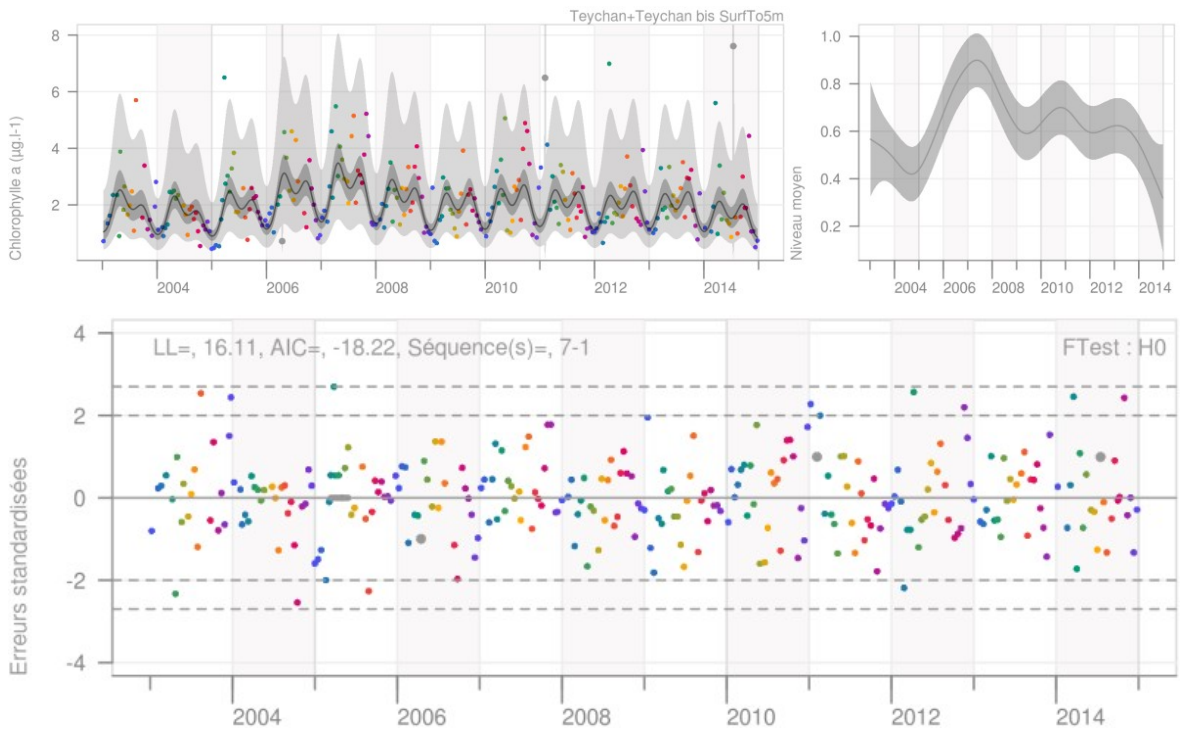
Teychan bis

1. Le premier modèle estime la variance d'observation à 0.200 log de chlorophylle a $\mu\text{g/L}$ et sa log-vraisemblance est de 4.31.



Trois valeurs exceptionnelles sont suggérées. Chronologiquement, la première est une valeur faible du printemps 2006 : 0.72 $\mu\text{g/L}$, comprise entre 2.3 et 4.57. La seconde est une concentration très élevée début 2011. La dernière est la plus forte concentration mesurée à plus de 7 $\mu\text{g/L}$.

2. Le modèle traitant ces trois valeurs comme exceptionnelles apporte un gain significatif de log-vraisemblance ($p=0.008$, $LL = 16.11$) et la variance d'observation diminue à 0.176 log de chlorophylle a $\mu\text{g/L}$.



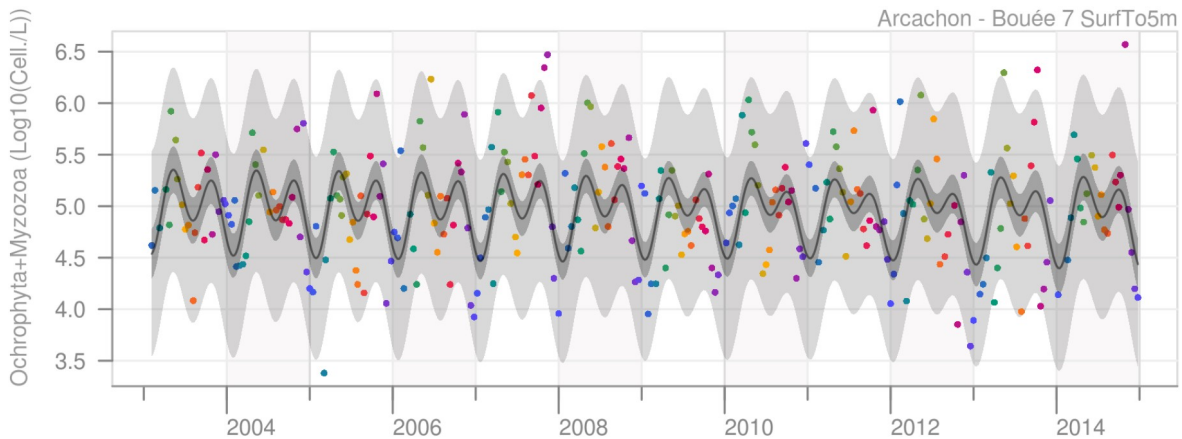
Ce dernier modèle ne suggère pas d'autre valeur exceptionnelle. La normalité des résidus standardisés est acquise aussi bien au regard du Q-Q plot, du test de Kolmogorov-Smirnov, que du Shapiro-Wilks et l'hypothèse de leur indépendance n'est pas rejetée par le test de Stoffer-Toloi. Toutefois, il présente une séquence de

7 résidus positifs au printemps 2005. En dépit de plusieurs essais d'interventions, les résultats obtenus se sont avérés moins satisfaisants, y compris d'un point de vue expertise biologique, écologique et terrain.

8.2 ABONDANCE PHYTOPLANCTONIQUE

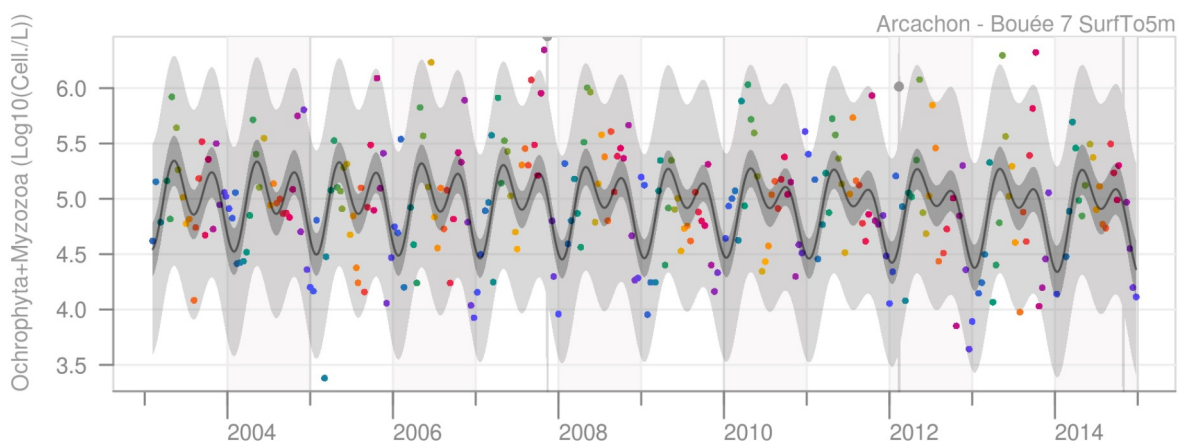
Bouée 7

1. Le premier modèle estime la variance d'observation à $0.243 \log_{10} \text{ cell./L}$ pour une log-vraisemblance de -14.41 .



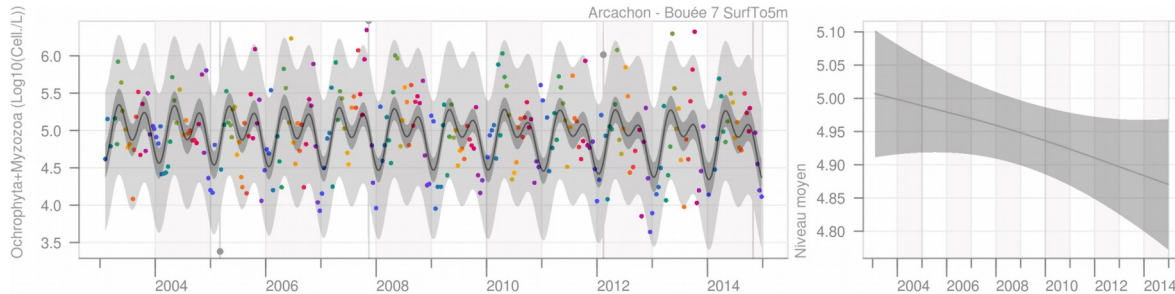
Il suggère trois valeurs exceptionnelles. Chronologiquement, la première fait partie d'un groupe de valeurs élevées entraînant une séquence d'erreurs standardisées positives dans la seconde moitié 2007 : c'est le même schéma que celui qui a amené pour la chlorophylle *a* à une intervention. La seconde est une valeur hivernale élevée, $6.01 \log_{10} \text{ cell./L}$, début 2012. Elle correspond à la même date que la seconde valeur exceptionnelle pour la chlorophylle *a*, le 16 février, qui est relevée comme un bloom dans le bulletin de la surveillance. Les valeurs voisines à plus et moins deux semaines sont 5.20 et 4.93, ce qui constitue sur un mois une multiplication puis une division d'un facteur 10 en cell./L. La dernière valeur au 31 octobre 2014 est un bloom à plus de 3 millions de cell./L composé à 96 % d'*Asterionnella glacialis*. La valeur correspondante de chlorophylle *a* est élevée mais dans l'intervalle de confiance à 95 % des observations. Pour l'abondance comme pour la biomasse, les valeurs voisines sont beaucoup plus en accord avec les valeurs saisonnières habituelles.

2. Le second modèle traite les trois valeurs exceptionnelles précédentes de manière appropriée pour une log-vraisemblance de -5.17 , gain significatif ($p < 0.001$). La variance d'observation est estimée à $0.221 \log_{10} \text{ cell./L}$.



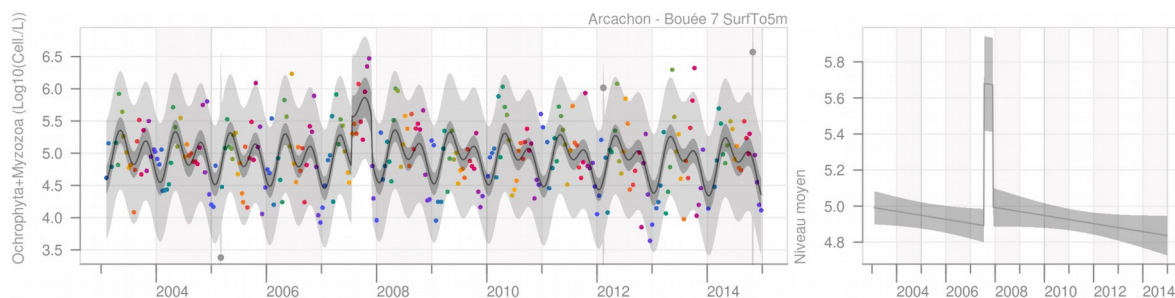
Il suggère une nouvelle valeur exceptionnelle début 2005, la plus faible concentration mesurée sur la période.

3. Le troisième modèle traite cette valeur comme exceptionnelle. La variance d'observation passe à $0.216 \log_{10} \text{ cell./L}$, log-vraisemblance à -2.32 , et le gain est significatif ($p = 0.017$).



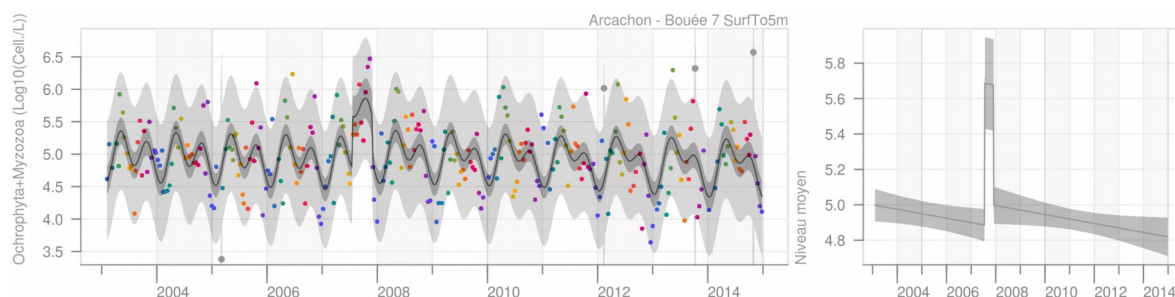
En 2007, du 25/07/2007 au 4/12/2007, 8 erreurs standardisées sont positives et une est négative et très proche de 0. Ainsi elles ne constituent pas une séquence de plus de 7 erreurs standardisées de même signe, mais correspondent à l'événement traité par intervention du 26/06/2007 au 4/12/2007 pour la chlorophylle *a*. Nous avons jugé que de manière identique, il y avait matière à suggérer une intervention de changement de niveau.

4. Après application, la log-vraisemblance du nouveau modèle est de -0.46 et la variance d'observation de $0.210 \log_{10} \text{ cell./L}$. Comme les modèles ne sont pas emboîtés on ne peut pas tester la significativité du gain de log-vraisemblance. Toutefois, l'indice d'Akaike est plus petit que celui du modèle précédent.



Une nouvelle valeur exceptionnelle est suggérée à l'automne 2013.

5. Le dernier modèle inclut le traitement de cette valeur exceptionnelle. La variance d'observation est de $0.204 \log_{10} \text{ cell./L}$ et la log-vraisemblance est de 2.22 . Le gain de log-vraisemblance est significatif ($p=0.021$).

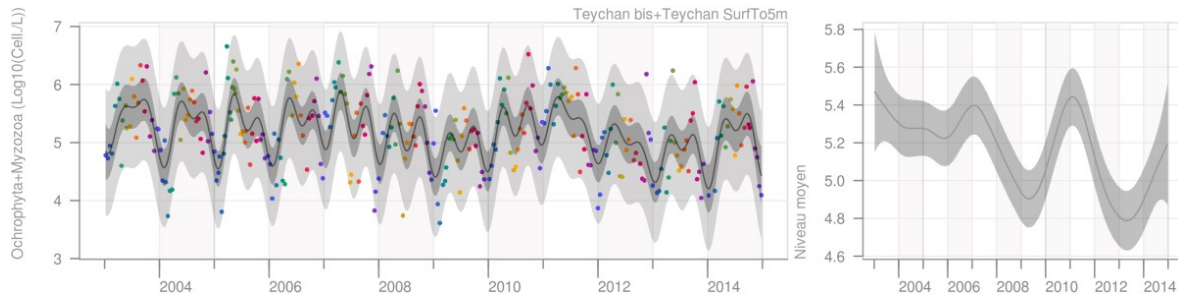


Ce modèle ne présente pas de séquence d'erreurs standardisées de même signe. La normalité des résidus standardisés est vérifiée aussi bien au regard du Q-Q plot, du test de Kolmogorov-Smirnov, que du Shapiro-Wilks et l'hypothèse de leur

indépendance n'est pas rejetée par le test de Stoffer-Toloi. Le modèle en l'état est reconnu comme pertinent par les experts biologique, écologique et terrain.

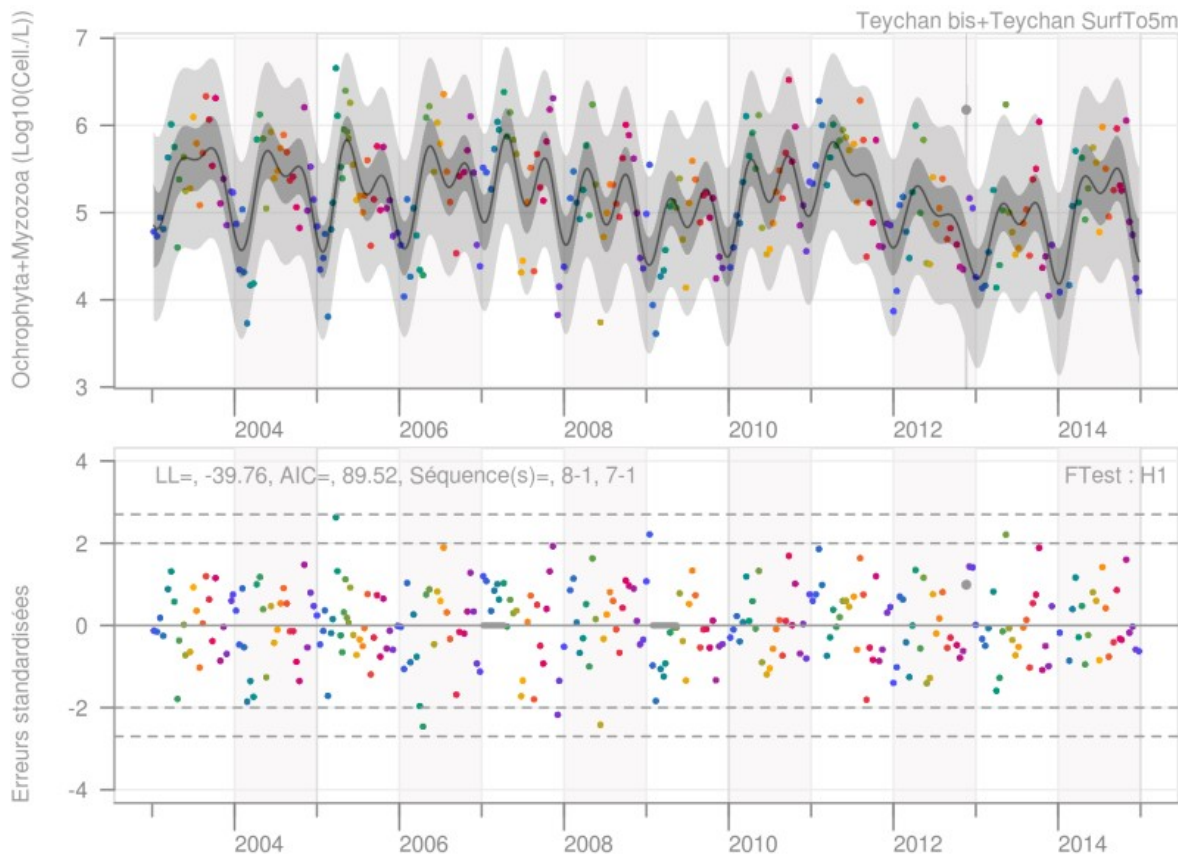
Teychan bis

1. Le premier modèle estime la variance d'observation à $0.251 \log_{10}$ cell./L avec une log-vraisemblance de -46.04 .



Une valeur exceptionnelle est suggérée fin 2012.

2. Lorsqu'elle est traitée comme telle, la variance d'observation diminue à $0.243 \log_{10}$ cell./L. La log-vraisemblance passe à -39.76 et le gain est significatif ($p < 0.001$).



Aucune autre valeur exceptionnelle ou changement de niveau n'est suggéré. Toutefois, les erreurs standardisées présentent deux séquences de signes identiques : l'une positive de la première mesure de 2007 jusqu'au 10 avril, l'autre négative début 2009. Pour 2007, les commentaires du bulletin relèvent :

En début d'année 2007, sur le point "Teychan bis", les abondances microphytoplanctoniques ont été supérieures à celles des 9 années précédentes. Le pic de floraison printanière s'est produit en avril. Des niveaux d'abondance très élevés ont été observés en mars, avril. De même,

le bloom automnal a été très marqué en novembre. Des abondances exceptionnellement basses ont été mesurées en juillet-août et décembre.

Pour 2009 :

Durant presque toute l'année 2009, sauf en janvier, les abondances microphytoplanctoniques ont été très inférieures à celles des 10 années précédentes. Les maxima ont été observés en janvier et durant l'été. Fin juillet, les abondances atteignaient à peine 400 000 cellules/L. L'année 2009 est donc caractérisée par une pauvreté phytoplanctonique très marquée, sans floraison printanière notable. Au regard de la série historique dont on dispose, cette situation n'avait pas été observée depuis 1993.

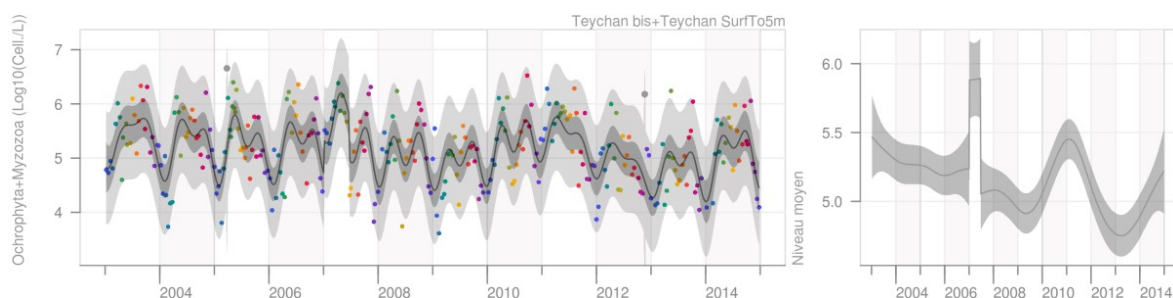
D'un point de vue météorologique, pour 2007 :

La France a connu un hiver exceptionnellement chaud, se situant au second rang des hivers les plus chauds observés depuis 1900. [...] Le printemps 2007 a été exceptionnellement chaud en France, se situant au premier rang des printemps les plus chauds observés depuis 1900.

L'année 2009^k compte parmi les événements météorologiques majeurs :

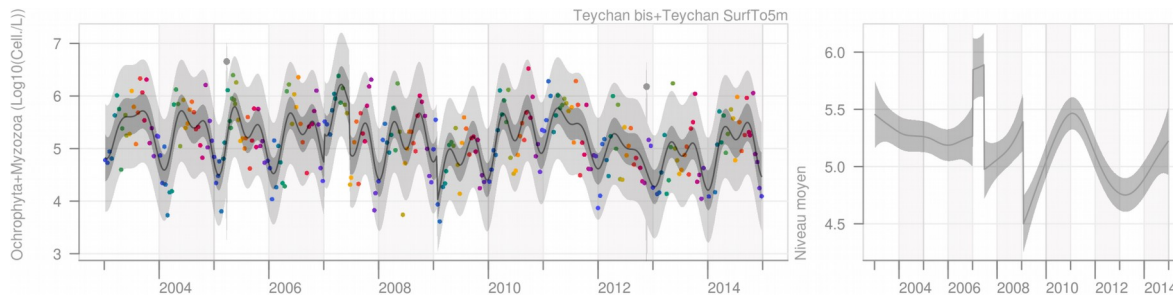
Tempête Klaus du 24 janvier 2009 : Une tempête exceptionnelle a touché le sud-ouest du pays le 24 janvier 2009.

3. Une intervention sur le niveau au 12 janvier 2007 ne permet pas de dissoudre complètement la séquence d'erreurs positives. Bien que la séquence d'erreur du même signe prenne fin au 10 avril, elle est suivie d'une erreur très proche de 0 puis de 3 erreurs de nouveau positives jusqu'au 26 juin. Une seconde intervention à cette dernière date entraîne la disparition de la séquence, mais le modèle résultant présente un gain de log-vraisemblance non significatif ($p = 0.081$). Toutefois, il fait apparaître une valeur exceptionnelle au printemps 2005, qui est la plus forte concentration observée sur l'intervalle 2003-2014. Lorsque cette valeur est traitée comme exceptionnelle, le modèle résultant présente un gain de log-vraisemblance significatif ($p = 0.008$) par rapport au second modèle et la variance d'observation est de $0.229 \log_{10} \text{ cell./L}$.



k. <http://www.meteofrance.fr/contenus-archives/bilan-2009/bilan-de-lannee-2009#>

4. Sur la base de ce modèle, une suggestion de changement de niveau au 3 février 2009, visant à traiter la séquence d'erreurs de même signe relevé au second modèle, entraîne un nouveau gain de log-vraisemblance ($p = 0.007$) et fait passer la variance d'observation à $0.221 \log_{10} \text{ cell./L}$.

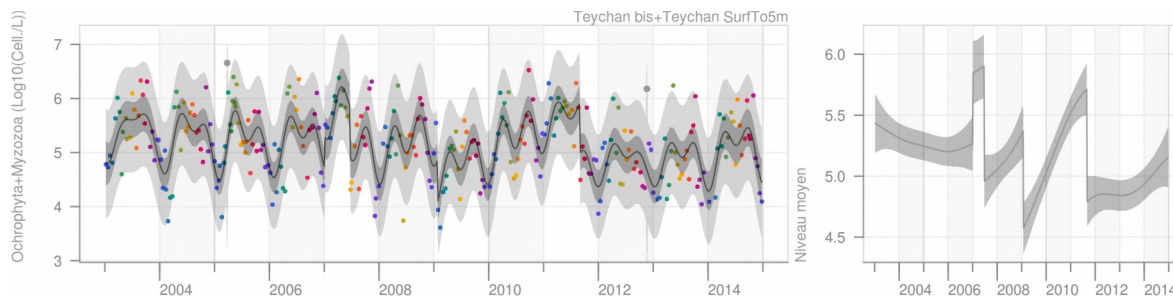


Ce modèle suggère un changement de niveau à la baisse fin 2011. D'un point de vue météorologique¹ :

Le mois de septembre a été particulièrement chaud : moyennée sur la France et l'ensemble du mois, la température a dépassé la normale de 2,1 °C. Toutefois, ces écarts de températures ont été relativement plus faibles sur la Bretagne et un peu plus marqués sur le Sud-Ouest.

Avec un déficit global à l'échelle du pays supérieur à 40 %, septembre 2011 s'inscrit parmi les mois de septembre remarquablement secs, sans être pour autant exceptionnel.

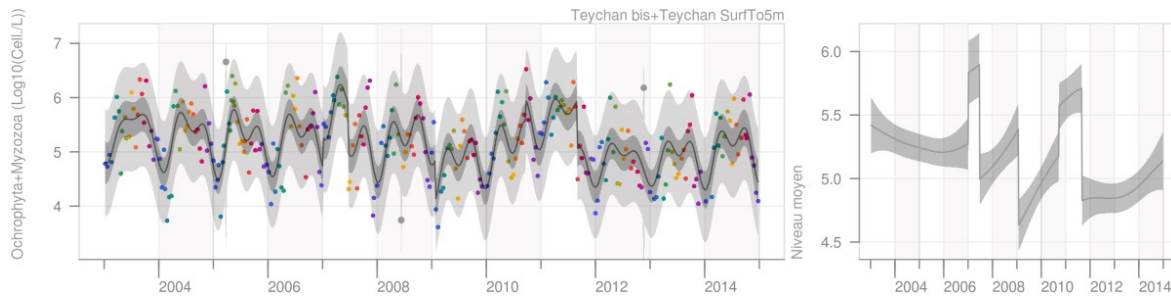
5. En ajoutant une intervention au 6 septembre 2011, date à laquelle les abondances sont réduites d'un facteur supérieur à 10 on obtient un modèle présentant un nouveau gain significatif de log-vraisemblance ($p = 0.003$, $LL = -24.03$) et une variance d'observation de $0.217 \log_{10} \text{ cell./L}$.



Ce nouveau modèle identifie un changement de niveau à l'automne 2010 et une valeur exceptionnelle le 12/06/2008.

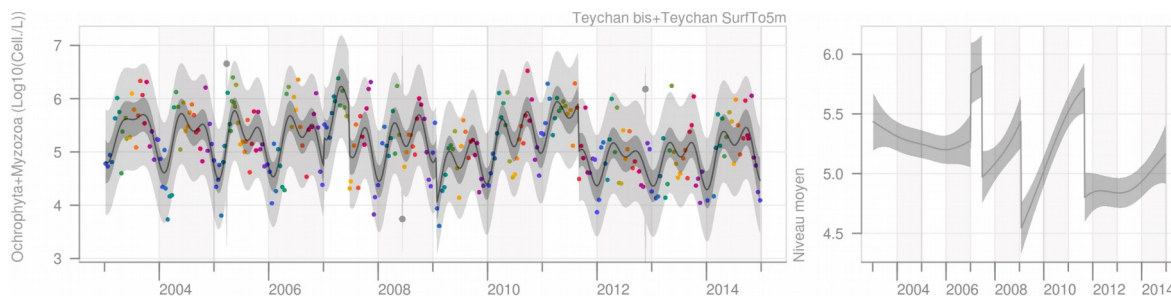
1. <http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/autres-annees/bilan-de-lannee-2011#>

6. En la traitant comme telle et en suggérant un changement le 30/09/2010 on obtient un dernier modèle dont le gain de log-vraisemblance est significatif ($p = 0.031$, $LL = -20.54$) et une variance d'observation de $0.211 \log_{10} \text{ cell./L}$.



D'un point de vue statistique, biologique et écologique, l'intervention en 2010 ne paraît pas essentielle : son influence est limitée et elle n'est pas associée à un commentaire particulier dans le bulletin de la surveillance non plus qu'à une caractéristique météorologique remarquable.

7. Ainsi, le modèle final n'inclut pas cette dernière intervention.



Ce dernier modèle ne présente aucune séquence d'erreurs du même signe, non plus que de valeurs exceptionnelles ou de changement de niveau. La variance d'observation est de $0.212 \log_{10} \text{ cell./L}$, la log-vraisemblance égale à -21.58 et est non significativement différente de celle du modèle précédent ($p = 0.851$). La normalité des résidus standardisés est vérifiée aussi bien au regard du Q-Q plot et du test de Kolmogorov-Smirnov et l'hypothèse de leur indépendance n'est pas rejetée par le test de Stoffer-Toloi.