

DIRECTION DE L'ENVIRONNEMENT ET DE L'AMENAGEMENT LITTORAL

Application de modèles dynamiques bayésiens à la
prévision des efflorescences à *Dinophysis* à Antifer
(baie de Seine)

par Dominique SOUDANT

D.E.A. de Biomathématiques
de l'université Paris VII,
soutenu le 9 septembre 1993 à
l'hôpital St Antoine (Paris).



R.INT.DEL/93.10/NANTES

DIRECTION DE L'ENVIRONNEMENT ET DE L'AMENAGEMENT LITTORAL

Application de modèles dynamiques bayésiens à la
prévision des efflorescences à *Dinophysis* à Antifer
(baie de Seine)

par Dominique SOUDANT

D.E.A. de Biomathématiques
de l'université Paris VII,
soutenu le 9 septembre 1993 à
l'hôpital St Antoine (Paris).

Direction de l'Environnement
Littoral

Service Qualité du Milieu

AUTEURS Dominique SOUDANT	CODE : N° R.INT. DEL/93.10/ NANTES
TITRE Application de modèles dynamiques bayésiens à la prévision des efflorescences à <i>Dinophysis</i> à Antifer (baie de Seine)	date : 14 septembre 93 tirage n° : 90 Nb pages : 30 Nb figures : 5 Nb photos :
CONTRAT (intitulé) N° _____	DIFFUSION libre <input checked="" type="checkbox"/> restreinte <input type="checkbox"/> confidentielle <input type="checkbox"/>

L'objectif de ce travail était de mettre en place un outil mathématique permettant de prévoir les concentrations en *Dinophysis cf. acuminata*, micro-algue toxique. Nous avons décidé d'utiliser des méthodes dynamiques de prévision bayésienne : modèles polynomiaux du premier, second et troisième ordre, modèle polynomial du premier ordre avec intervention, et modèle de régression multilinéaire.

Les modèles polynomiaux du premier, second et troisième ordre se sont avérés inadaptés. Le modèle de régression multilinéaire a été testé avec les variables concentrations en phosphate (PO₄) et en nitrate (NO₃). Ce modèle est insatisfaisant du fait de l'inconstance des relations entre la concentration en *Dinophysis cf. acuminata* et les régresseurs. Le modèle polynomial du premier ordre avec intervention, testé avec les données de la plage Benoît en baie du Pouliguen, donne de bons résultats qualitatifs, mais commet parfois d'importantes erreurs quantitatives.

Les résultats obtenus sont insatisfaisants du point de vue de l'objectif que nous nous étions fixé. Mais les modèles dynamiques bayésiens, du fait de leurs comportements, restent des méthodes encourageantes. Il existe des modèles permettant de modéliser des processus issus de lois de la famille exponentielle (*exponential family dynamic models*), ainsi que des modèles plus complexes (*multi-process models, mixture models*). Ces raffinements ne peuvent masquer le manque de connaissances fondamentales concernant la micro-algue (mode de nutrition, variabilités spatiales et temporelles).

Mots-clés : Modèle dynamique bayésien - Prévision - Séries chronologiques - *Dinophysis* - Antifer.

Tables des matières

AVANT PROPOS	2
1 Introduction.....	3
2 Données et méthodes	4
2.1 Données.....	4
2.2 Méthodes	4
2.2.1 Modèle dynamique linéaire général.....	4
2.2.2 Modèle dynamique linéaire polynomial général.....	5
2.2.3 Modèle dynamique linéaire polynomial du premier ordre.....	6
2.2.4 Modèles dynamiques linéaires polynomiaux du second et du troisième ordre.....	6
2.2.5 Modèle dynamique de régression linéaire multivariée	8
2.2.6 Limites et extensions	8
3 Résultats 11	
3.1 Modèles polynomiaux	11
3.2 Modèle linéaire dynamique polynomial du premier ordre avec intervention.....	13
3.3 Modèle dynamique de régression linéaire multivariée.....	14
4 Discussion.....	15
5 Conclusion	17
6 Bibliographie	18
TABLE DES ANNEXES	21

AVANT PROPOS

Ce travail a été réalisé au sein du service DEL/QM du centre IFREMER de Nantes.

J'exprime toute ma gratitude à Monsieur Benoît Beliaeff, pour avoir toujours su être disponible malgré ses multiples occupations.

Je remercie vivement Monsieur Guy Thomas pour ces précieux conseils.

Je témoigne de mon amitié à tous les membres du centre IFREMER de Nantes pour leur gentillesse et leur disponibilité.

1 Introduction

En 1983, plusieurs épidémies de diarrhées ont été imputées à la consommation de coquillages. Depuis, cette toxicité a été associée à la présence en milieu marin de dinoflagellés, algues unicellulaires, du genre *Dinophysis cf acuminata*¹ [1, 2]. La toxine responsable des intoxications, l'acide okadaïque, a été retrouvé à la fois dans les moules et dans les cellules de *Dinophysis* [3, 4, 5]. En 1984, l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) a mis en place le Réseau de surveillance Phytoplanctonique (REPHY). Ce réseau couvre les côtes françaises avec l'aide des laboratoires côtiers. Ces derniers effectuent de mai à septembre des prélèvements d'eau de mer et de coquillages sur 110 points d'alerte. Lorsque la quantité de *Dinophysis* dépasse le seuil empirique de 200 cellules par litre, les "tests-souris" sont effectués à partir des hépatopancreas des coquillages. Un test positif conduit à l'arrêt de la commercialisation des coquillages. L'interdiction est levée après deux tests négatifs.

Le *Dinophysis* a fait l'objet de nombreuses recherches [6, 7, 8, 9, 10, 11]. Il apparaît d'abord au large en relation avec une stratification thermo-haline [12, 13]. Il reste que la micro-algue n'a jamais pu être cultivée en laboratoire, les facteurs environnementaux ou biologiques de déclenchement des efflorescences sont mal cernés et la cinétique de contamination des coquillages n'a jamais pu être observée ni expérimentalement, ni *in situ*. Des approches statistiques utilisant des méthodes descriptives, la corrélation simple et multiple, la régression multivariée linéaire et non linéaire, l'analyse discriminante et l'analyse de données se sont avérées inadaptées [14]. L'utilisation de modèles ARIMA, de modèles d'espace d'états et du filtre de Kalman ont donné de bons résultats mais présentent des inconvénients opérationnels (fréquence de prélèvement régulière) au niveau de la prédiction [15]. Enfin une modélisation en vue de simulation a été réalisée [16]. Ce type de modèle a permis de tester des hypothèses et d'en émettre d'autres. Mais ses capacités prédictives sont faibles et certaines caractéristiques (conditions initiales, paramètres, etc.) ne sont pas vérifiables.

Notre travail constitue une première approche de formalisation pour le déclenchement de l'alerte phytoplanctonique. Notre objectif est de mettre en place un outil mathématique permettant de prévoir la dynamique de la concentration en *Dinophysis*. En particulier nous insisterons sur la nécessité de pouvoir détecter les efflorescences algales le plus tôt possible. La méthode de prévision à court terme présentée par Harrison et Stevens [17, 18, 19, 20] permet de surveiller en temps réel l'évolution d'un processus. Nous nous sommes limités aux modèles linéaires dynamiques polynomiaux du premier, du second et du troisième ordre, et au

¹Par la suite nous ferons référence à *Dinophysis*.

modèle de régression linéaire multivarié dynamique. Les résultats que nous obtiendrons avec ces modèles fourniront une base de réflexion sur l'adéquation de la stratégie d'alerte mise en oeuvre avec les objectifs de protection de la santé humaine. De plus les méthodes utilisées pourront être appliquées à d'autres problèmes de surveillance du milieu maritime, pour d'autres algues toxiques (e.g. *Alexandrium minutum*) ou pour la microbiologie (e.g. coliformes fécaux).

2 Données et méthodes

2.1 Données

De 1987 à 1992, le laboratoire municipal du Havre, sous contrat avec l'IFREMER, a recueilli au site d'Antifer (Seine Maritime) des données journalières. Les prélèvements ont en général été effectués du premier juin au 30 septembre. Nous avons décidé d'utiliser la base de données de 1988, année où le nombre de variables mesurées a été le plus important, et de 1987 où la série présente un nombre important de pics.

Nous avons également décidé d'utiliser des données de microbiologie recueillies à la plage Benoît en baie du Pouliguen (Loire-Atlantique). Nous avons choisi ces données car en ce lieu il a été montré qu'il existait une relation entre les précipitations et les taux de coliformes fécaux [21].

Ces variables et les détails s'y rapportant sont donnés en ANNEXE I.

2.2 Méthodes

Les méthodes utilisées sont celles décrite dans l'ouvrage de West [20]. Les modèles utilisés sont présentés sous leur forme à variances constantes. Par convention, les lettres soulignées représenteront des vecteurs ou des matrices.

2.2.1 Modèle dynamique linéaire général

Dans la formalisation adoptée par West et Harrison, le modèle dynamique linéaire général est représenté par le quadruplet $\{\underline{F}_t, \underline{G}_t, \underline{V}_t, \underline{W}_t\}$ où pour tout t :

- (a) \underline{F}_t est une matrice $(n \times r)$ connue,
- (b) \underline{G}_t est une matrice $(n \times n)$ connue,

- (c) \underline{V}_t est une matrice ($r \times r$) connue,
 (d) \underline{W}_t est une matrice ($n \times n$) connue.

Le modèle a pour forme :

$$\begin{aligned} \underline{Y}_t &= \underline{F}_t' \underline{\theta}_t + \underline{v}_t & \underline{v}_t &\rightarrow N(0, \underline{V}_t) \\ \underline{\theta}_t &= \underline{G}_t \underline{\theta}_{t-1} + \underline{\omega}_t & \underline{\omega}_t &\rightarrow N(0, \underline{W}_t) \end{aligned}$$

où \underline{Y}_t est le vecteur ($r \times 1$) des séries temporelles étudiées et $\underline{\theta}_t$ le vecteur ($n \times 1$) des paramètres. On peut également écrire :

$$\begin{aligned} (\underline{Y}_t | \underline{\theta}_t) &\rightarrow N(\underline{F}_t' \underline{\theta}_t, \underline{V}_t) \\ (\underline{\theta}_t | \underline{\theta}_{t-1}) &\rightarrow N(\underline{G}_t \underline{\theta}_{t-1}, \underline{W}_t) \end{aligned}$$

Ces distributions sont également conditionnelles à D_{t-1} , ensemble des informations *a priori* disponibles au temps t . Ceci inclut les variances \underline{W}_{t-1} et \underline{V}_{t-1} , les observations passées $\underline{Y}_{t-1}, \underline{Y}_{t-2}, \dots$, ainsi que les ensembles D_n , pour $n = 0, 1, \dots, t-1$. De plus, chacune des séquences \underline{v}_t et $\underline{\omega}_t$ est indépendante et elles sont mutuellement indépendantes.

2.2.2 Modèle dynamique linéaire polynomial général

Ce type de modèle est représenté par le quadruplet $\{\underline{E}_n, \underline{L}_n, \underline{V}_t, \underline{W}_t\}$, où :

$$\underline{E}_n = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{et} \quad \underline{L}_n = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Ces modèles décomposent l'observation en une série de "constantes dynamiques". Dans les faits on utilise surtout les modèles correspondant aux trois premiers ordres.

2.2.3 Modèle dynamique linéaire polynomial du premier ordre

Le modèle du premier ordre est représenté par le quadruplet $\{1, 1, V_t, W_t\}$. Les équations sont les suivantes :

équation d'observation	$Y_t = \mu_t + v_t$	$v_t \rightarrow N(0, V_t)$ $V_t = V$
équation du système	$\mu_t = \mu_{t-1} + \omega_t$	$\omega_t \rightarrow N(0, W_t)$ $W_t = W$
information	$(\mu_{t-1} D_{t-1}) \rightarrow N(m_{t-1}, C_{t-1})$ $(\mu_t D_{t-1}) \rightarrow N(m_t, R_t)$	$R_t = C_{t-1} + W$
prévision	$(Y_t D_{t-1}) \rightarrow N(f_t, Q_t)$	$f_t = m_{t-1}$ $Q_t = R_t + V_t$
information initiale	m_0, C_0, W, V	
équations de mise à jour	$m_t = m_{t-1} + A_t e_t$ $C_t = A_t V$ où $e_t = Y_t - f_t$ et $A_t = \frac{R_t}{Q_t}$	

2.2.4 Modèles dynamiques linéaires polynomiaux du second et du troisième ordre

Le modèle du second ordre est dit à croissance linéaire et celui du troisième ordre à croissance quadratique. Nous ne présenterons ici que les équations pour le modèle du second ordre. Ce type de modèle peut être formalisé par le quadruplet :

$$\{\underline{E}_2, \underline{L}_2, V_t, W_t\} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, V_t, W_t \right\}.$$

Les équations sont :

équation d'observation	$Y_t = \mu_t + v_t$	$v_t \rightarrow N(0, V_t)$ $V_t = V$
------------------------	---------------------	--

équation du système

$$\begin{aligned}\mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{1,t} & \omega_{1,t} &= \omega_1 \\ \beta_t &= \beta_{t-1} + \omega_{2,t} & \omega_{2,t} &= \omega_2\end{aligned}$$

$$\underline{\omega}_t = (\omega_{1,t}, \omega_{2,t}) \quad \underline{\omega}_t \rightarrow N(0, \underline{W}_t)$$

$$\underline{W}_t = W = \underline{L}_2 \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} \underline{L}_2' = \begin{pmatrix} W_1 & W_3 \\ W_3 & W_2 \end{pmatrix}$$

information

$$\begin{aligned}(\underline{\theta}_{t-1} | D_{t-1}) &\rightarrow N(\underline{m}_{t-1}, C_{t-1}) & \underline{m}_{t-1} &= \begin{pmatrix} m_{t-1} \\ b_{t-1} \end{pmatrix} \\ (\underline{\theta}_t | D_{t-1}) &\rightarrow N(\underline{a}_t, R_t) & \underline{a}_t &= \underline{L}_2 \underline{m}_{t-1} \\ & & R_{t-1} &= \underline{L}_2 C_{t-1} \underline{L}_2' + W\end{aligned}$$

prévision

$$\begin{aligned}(Y_t | D_{t-1}) &\rightarrow N(f_t, Q_t) & f_t &= \underline{E}_2' \underline{a}_t \\ & & Q_t &= \underline{E}_2' R_t \underline{E}_2 + V\end{aligned}$$

information initiale

$$m_0, C_0, W, V$$

équations de mise à jour

$$\begin{aligned}\underline{m}_t &= \underline{a}_t + \underline{A}_t e_t \\ C_t &= R_t - \underline{A}_t \underline{A}_t' Q_t \\ \text{où } e_t &= Y_t - f_t \text{ et } \underline{A}_t = R_t \underline{E}_2 Q_t^{-1}\end{aligned}$$

Il existe par ailleurs des relations bornant les éléments du vecteur \underline{A}_t .

Dans ce modèle, μ_t représente le niveau moyen de la série et β_t représente l'augmentation, ou bien encore la pente de la série entre deux temps. Pour le modèle du troisième ordre le paramètre supplémentaire, γ_t représente le changement dans la pente.

2.2.5 Modèle dynamique de régression linéaire multivariée

Le quadruplet de ce modèle est :

$$\{\underline{F}_t, \underline{I}, \underline{V}_t, \underline{W}_t\} = \left\{ \begin{pmatrix} 1 \\ X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{n,t} \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \underline{V}_t, \underline{W}_t \right\}.$$

Les équations de ce modèle sont les suivantes :

équation d'observation	$\underline{Y}_t = \underline{F}_t' \underline{\theta}_t + \underline{v}_t$	$\underline{v}_t \rightarrow N(0, \underline{V}_t)$ $\underline{V}_t = \underline{V}$
équation du système	$\underline{\theta}_t = \underline{\theta}_{t-1} + \underline{\omega}_t$	$\underline{\omega}_t \rightarrow N(0, \underline{W}_t)$
information	$(\underline{\theta}_{t-1} D_{t-1}) \rightarrow N(\underline{m}_{t-1}, \underline{C}_{t-1})$ $(\underline{\theta}_t D_{t-1}) \rightarrow N(\underline{a}_t, \underline{R}_t)$	$\underline{a}_t = \underline{m}_{t-1}$ $\underline{R}_{t-1} = \underline{C}_{t-1} + \underline{W}$
prévision	$(Y_t D_{t-1}) \rightarrow N(f_t, Q_t)$	$f_t = \underline{F}_t' \underline{a}_t$ $Q_t = \underline{F}_t' \underline{R}_t \underline{F}_t + \underline{V}$
information initiale	$\underline{m}_0, \underline{C}_0, \underline{W}, \underline{V}$	
équations de mise à jour	$\underline{m}_t = \underline{a}_t + \underline{A}_t e_t$ $\underline{C}_t = \underline{R}_t - \underline{A}_t \underline{A}_t' Q_t$ où $e_t = Y_t - f_t$ et $\underline{A}_t = \underline{R}_t \underline{F}_t Q_t^{-1}$	

2.2.6 Limites et extensions

Comme nous venons de le voir, dans tous les cas ces méthodes requièrent une connaissance *a priori* du phénomène étudié. En particulier, il est nécessaire d'avoir une idée de la variabilité du processus et de ses composantes. Cependant, étant donné que les variances

sont constantes, ces modèles convergent rapidement. Ainsi, pour le modèle dynamique linéaire polynomial du premier ordre on obtient pour les quantités suivantes en posant $r = \frac{W}{V}$:

$$(i) : A_t \rightarrow r(\sqrt{1+4/r}-1)/2$$

$$(ii) : C_t \rightarrow AV$$

$$(iii) : R_t \rightarrow \frac{C}{1-A}$$

$$(iv) : Q_t \rightarrow \frac{V}{1-A}$$

Nous avons remarqué que la quantité A_t , désignée sous le terme coefficient adaptatif par les auteurs de ces méthodes, présente un caractère adaptatif fugace. En effet, moins d'une dizaine d'itérations aboutissent à la convergence de ces quantités. Ces convergences sont dépendantes du rapport des variances des modèles. Ceci a pour résultat, à la limite (au sens mathématique), que la prévision au pas suivant n'est autre que la prévision au pas courant plus une proportion fixe de l'erreur commise au pas courant. De plus, pour deux modèles M et M' différant uniquement par les variances, mais de rapport r identique, les prédictions sont les mêmes.

Dans le cas de variances inconnues et constantes les auteurs proposent les modifications suivantes :

$$(i) : \delta = 1 - A \quad W_t = C_{t-1} \frac{(1-\delta)}{\delta}$$

$$R_t = C_{t-1} + W_t = \frac{C_{t-1}}{\delta}$$

$$(ii) : \phi = V^{-1} \quad (\phi | D_{t-1}) \rightarrow G(n_{t-1}/2, d_{t-1}/2)$$

$$E(\phi | D_{t-1}) = \frac{n_{t-1}}{d_{t-1}}$$

$$S_{t-1} = \frac{d_{t-1}}{n_{t-1}}$$

δ est appelé "facteur d'escompte" (*discount factor*) et ϕ est la précision. Ainsi dans le cas du modèle dynamique linéaire du premier ordre, ces modifications entraînent :

$$\text{équation d'observation} \quad Y_t = \mu_t + v_t \quad v_t \rightarrow N(0, V_t)$$

$$\text{équation du système} \quad \mu_t = \mu_{t-1} + \omega_t \quad \omega_t \rightarrow T_{n_{t-1}}(0, W_t)$$

information	$(\mu_{t-1} D_{t-1}) \rightarrow T_{n_{t-1}}(m_{t-1}, C_{t-1})$ $(\mu_t D_{t-1}) \rightarrow T_{n_{t-1}}(m_{t-1}, R_t)$	$R_t = C_{t-1}/\delta$
prévision	$(Y_t D_{t-1}) \rightarrow T_{n_{t-1}}(f_t, Q_t)$	$f_t = m_{t-1}$ $Q_t = R_t + S_{t-1}$
information initiale	$m_0, C_0, d_0, n_0, \delta$	
équations de mise à jour	$m_t = m_{t-1} + A_t e_t$ $C_t = (S_t/S_{t-1})(R_t - A_t^2 Q_t)$ $n_t = n_{t-1} + 1$ $d_t = d_{t-1} + \frac{S_{t-1} e_t^2}{Q_t}$ où $e_t = Y_t - f_t$ et $A_t = \frac{R_t}{Q_t}$	

Pour structurer les séquences des variances \underline{W}_t et \underline{V}_t les auteurs conseillent d'utiliser des relations connues. Nous ne disposons pas de telles informations.

Enfin, un des avantages de ces méthodes est qu'elles sont ouvertes à l'utilisateur par le biais de l'intervention. C'est à dire que si l'on sait qu'un événement va influencer la série temporelle, on peut modifier la valeur et la variance de la prévision. Cette intervention a pour conséquence une modification des quantités du modèle et en particulier du coefficient adaptatif. Nous avons donc décidé de chercher une variable ou une combinaison de variables permettant d'anticiper une modification brutale de la série temporelle. Pour cela, nous avons décidé de choisir l'information utilisée pour cette intervention à partir des concentrations en *Dinophysis* et des variables ou combinaison de variables, sachant que cette information doit être décalée temporellement par rapport au taux de la micro-algue.

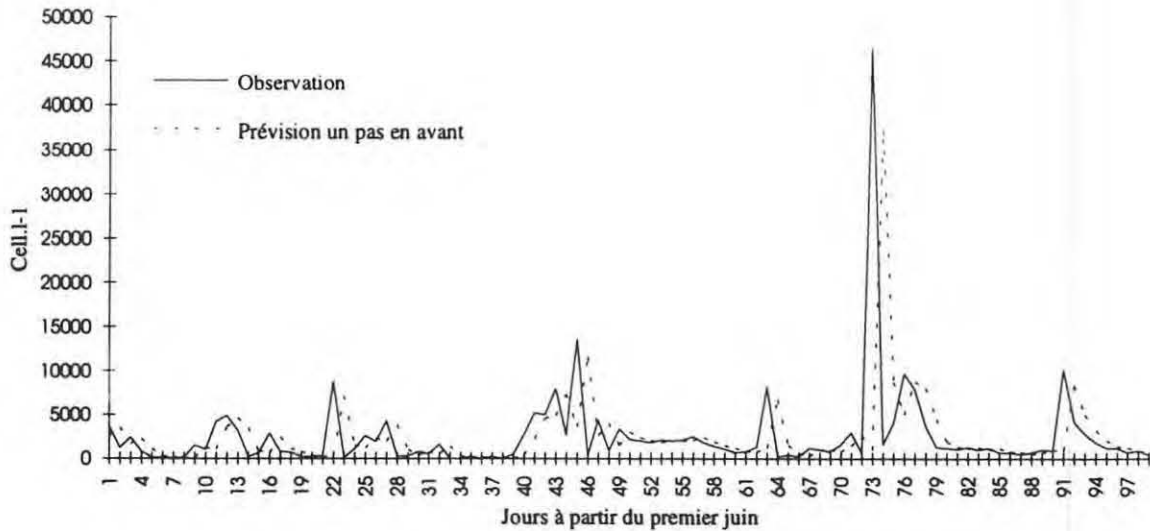


Figure 1 : Observation et prévision des concentrations en *Dinophysis* par le modèle dynamique linéaire polynomial du premier ordre à Antifer (1987).

3 Résultats

3.1 Modèles polynomiaux

Pour ces modèles nous avons utilisé la série de données de 1987. Nous avons décidé de fixer arbitrairement les paramètres de ces modèles en fonction des coefficients adaptatifs et de la variance de la prévision à la limite. On peut voir sur la série observée en 1987 que les valeurs semblent être réparties autour de 2500 plus ou moins 2500. Dans l'hypothèse de normalité du processus étudié, nous avons décidé de fixer la variance de la prévision après convergence à $\left(\frac{2500}{2}\right)^2$. Pour les coefficients adaptatifs nous avons choisi 0,8 pour A_{μ} , 0,5 pour A_{β} , et 0,1 pour A_{γ} , ces valeurs respectant, quelque soit l'ordre du modèle, les relations qui bornent ces quantités. Ces coefficients relativement élevés permettront de toujours prendre en compte, même après convergence, une part importante de l'erreur de prédiction commise. Des coefficients plus faibles n'amèneraient qu'à un lissage de la série. Les

conditions initiales sont reportées en ANNEXE II. Les résultats de ces modèles sont donnés dans les figures 1 à 3.

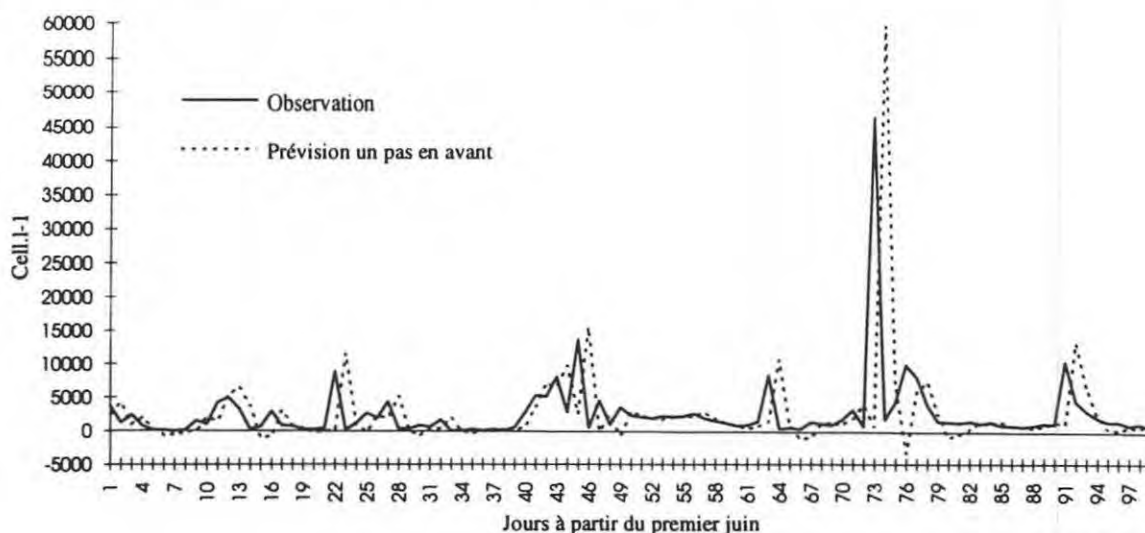


Figure 2 : Observation et prévision des concentrations en *Dinophysis* par le modèle dynamique linéaire polynomial du second ordre à Antifer (1987).

On remarque que pour les modèles d'ordre 2 et 3 certaines prévisions sont négatives. Nous avons décidé de conserver ces valeurs car ramener ces prévisions à la valeur 0 implique que l'on réalise une intervention, ce qui n'est pas ici notre objet. De plus l'erreur commise serait alors plus faible, ce qui implique que la correction de l'erreur serait également réduite.

On constate dans les trois cas que les modèles ne prévoient pas les augmentations brutales des taux de *Dinophysis*. Donc, sans information extérieure à la série temporelle étudiée on ne peut pas construire un modèle respectant les objectifs que nous nous sommes fixés. Il apparaît donc nécessaire, si l'on veut utiliser ces modèles, de pratiquer l'intervention. Pour qu'un modèle avec intervention soit efficace, l'information utilisée doit être disponible avant l'observation et significative d'un changement dans la série. De plus, elle doit être en phase, le plus souvent possible avec les augmentations brutales. Ces qualités doivent en outre être reproduites sur plusieurs années. Les variables dont nous disposions ne semblaient réunir ces propriétés. C'est pourquoi nous avons choisi les données de la plage Benoît afin d'illustrer le modèle linéaire dynamique polynomial du premier ordre avec intervention.

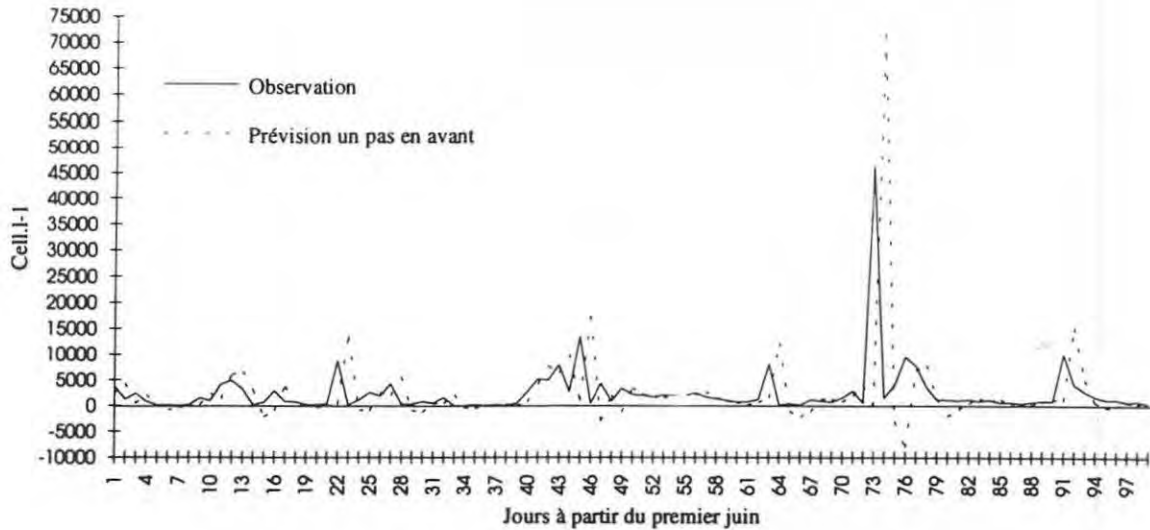


Figure 3 : Observation et prévision des concentrations en *Dinophysis* par le modèle dynamique linéaire polynomial du troisième ordre à Antifer (1987).

3.2 Modèle linéaire dynamique polynomial du premier ordre avec intervention.

La série étudiée est le taux de coliformes fécaux mesuré dans les coquillages, et la variable d'intervention est la somme cumulée des précipitations des quatre jours précédents. Nous avons choisi de fixer le coefficient adaptatif à 0,5. Les variances ont été spécifiées de sorte qu'à sa limite la variance de la prévision soit égale à la variance de la série (ANNEXE II). Nous avons choisi d'intervenir lorsque les précipitations étaient supérieures à un seuil arbitraire de 40 millimètres. La forme de l'intervention est une multiplication de la moyenne et de la variance *a posteriori* par les précipitations divisées par 40. De cette façon le modèle n'est pas sensible aux faibles précipitations et l'intervention ne peut diminuer la prévision².

²Ces valeurs ont été les premières utilisées. D'autres essais ont été tentés mais n'ont donné que des résultats similaires ou inférieurs.

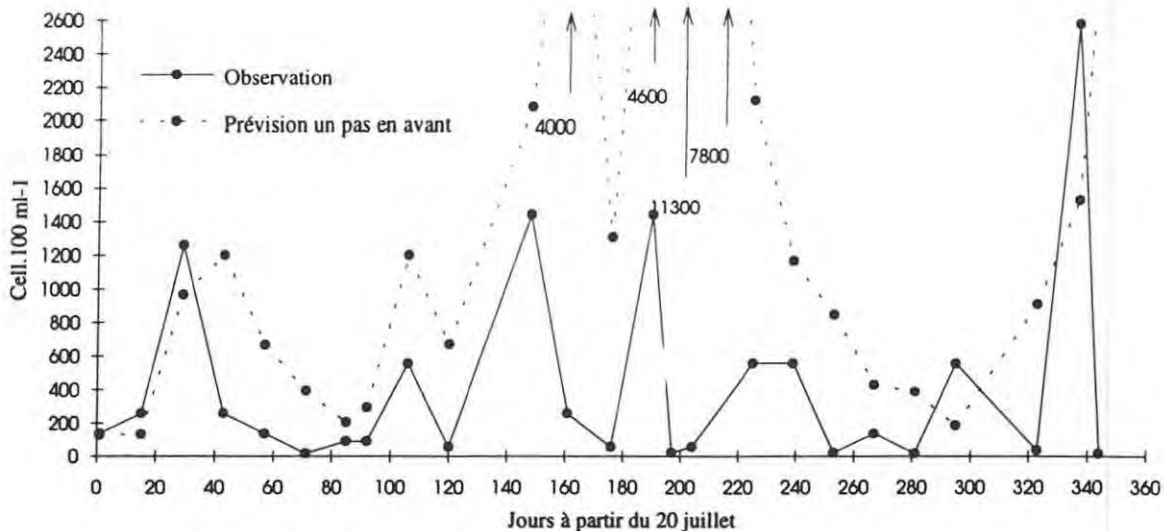


Figure 4 : Observation et prévision des taux de coliformes fécaux par le modèle dynamique linéaire polynomial du premier ordre avec intervention à la plage Benoît.

On remarque (figure 4) que qualitativement ce modèle suit approximativement l'évolution des taux de coliformes fécaux. Quantitativement il y a quelques prédictions correctes et quelques erreurs très importantes. En particulier, les augmentations des prévisions au centre de la figure sont dues à des successions de fortes valeurs de précipitations. Lors des augmentations, la pente de la prévision est à peu près équivalente à celle observée. Du fait de la grande variabilité des erreurs, l'information apportée par ce modèle est que s'il pleut beaucoup les quatre jours précédants le prélèvement, alors on risque d'observer des taux de coliformes fécaux élevés. Mais ce modèle est très simple.

3.3 Modèle dynamique de régression linéaire multivariée

De la même façon que pour l'intervention, les variables utilisées pour ce modèle doivent être disponibles la veille de l'observation. Nous avons choisi d'utiliser les concentrations en phosphates (PO_4) et en nitrates (NO_3), sels minéraux assimilables par le *Dinophysis*, la veille du prélèvement. De plus ces variables semblent liées aux augmentations brutales des taux de *Dinophysis*. Les paramètres du modèle ont été fixés à partir d'une régression linéaire classique utilisant ces mêmes variables. Ces valeurs sont données en ANNEXE II.

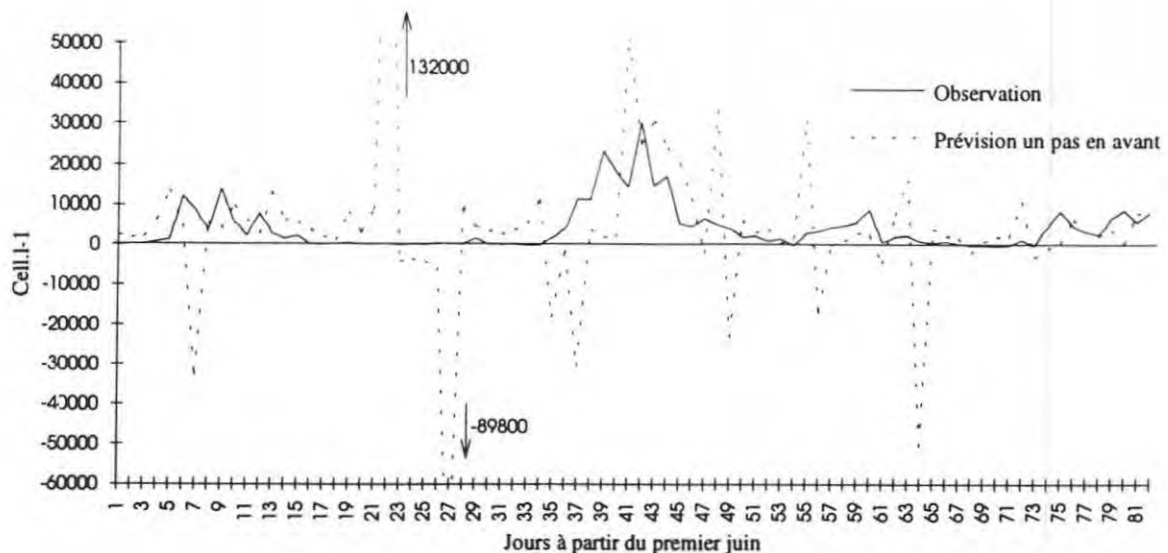


Figure 5 : Observation et prévision des concentrations en *Dinophysis* par le modèle dynamique de régression linéaire utilisant les concentrations en phosphates et en nitrates à Antifer (1988).

Comme on le voit sur la figure 5, ce modèle n'est pas satisfaisant. Ceci est dû au fait qu'il n'existe pas, même sur de courtes périodes, de relation linéaire stable entre les variables choisies et le *Dinophysis*. Le modèle ne cesse pas de se réajuster mais en vain.

4 Discussion

Tous les modèles que nous avons présentés sont insatisfaisants du point de vue des objectifs que nous nous étions fixés pour *Dinophysis*. Ceci peut être imputé aux modèles et aux données.

Comme nous l'avons vu ces modèles requièrent une connaissance *a priori* du processus étudié et en particulier des variances des composantes des modèles. Mais nous avons la possibilité d'utiliser le facteur d'escompte δ et la précision ϕ . Ces modifications ont été programmées pour tous les modèles. Toutes choses égales d'ailleurs, les prévisions obtenues avec ces modèles modifiés ne diffèrent pas des résultats qui ont été présentés. En effet, dans un cas on choisit les variances de manière à obtenir une valeur donnée du coefficient adaptatif après convergence, et dans l'autre la valeur du facteur d'escompte fixe la valeur limite de ce coefficient.

En ce qui concerne l'intervention, Harrison et West donnent un algorithme d'intervention automatique basé sur des calculs de vraisemblance. Mais en l'absence de variable adéquate cette possibilité nous reste toujours fermée.

Les modèles que nous avons utilisés reposent sur les hypothèses de normalité du processus et de l'existence d'une relation linéaire entre leurs composantes. La série dont nous disposons n'est pas nécessairement issue d'un processus normal. En général, les distributions phytoplanctoniques sont plutôt de type binomiale négative, log-normale ou poisson [e.g. 22, 23,]. On peut donc supposer que les modèles dynamiques non linéaires et les modèles dynamiques de la famille exponentielle auraient permis d'obtenir de meilleurs résultats. Enfin les modèles comportant plusieurs processus (*multi-process models*, *mixture models*) offrent d'intéressantes perspectives. Il s'agit d'une modélisation applicable à des séries chronologiques non linéaires, non normales et non stationnaires. En effet, on considère un processus M formé des processus M_1, M_2, \dots, M_n et un vecteur de n poids compris entre 0 et 1 et dont la somme fait 1 : la prévision de ce modèle est la somme pondérée par les n poids des n prévisions des modèles M_1, M_2, \dots, M_n . Le vecteur des pondérations est donc un vecteur des probabilités des modèles. Ce vecteur est de plus mis à jour après chaque observation. On trouvera en particulier dans l'article de Bolstad [24] un exemple d'application de ce type de modèle à la détection des pics d'hormones de croissance chez le veau.

Le réseau de surveillance phytoplanctonique est le destinataire final de ce travail. Nous soulèverons donc les faits suivants. Comme nous avons pu le constater, les concentrations en *Dinophysis* sont extrêmement variables d'un jour à l'autre. La fréquence hebdomadaire d'échantillonnage du réseau de surveillance n'est donc pas suffisante pour pouvoir détecter le seuil considéré comme critique de 200 cellules par litre. Par ailleurs, les comptages des cellules ne sont effectués qu'une seule fois par prélèvement et, du fait de la dilution utilisée, la quantité minimale détectable est de 100 cellules par litre. Il en va de même pour les données utilisées dans cette étude. On peut donc raisonnablement avancer qu'une part non négligeable de la variabilité observée est due aux méthodes d'échantillonnage et de comptage. Il s'ensuit qu'une solution permettant de déclencher l'alerte de façon plus adaptée serait d'effectuer plusieurs mesures par prélèvement et de multiplier la fréquence d'échantillonnage. Dans ce cas, il ne serait sans doute plus nécessaire de recourir à un modèle prédictif. Mais les coûts qu'entraînerait une telle stratégie seraient rédhibitoires.

Enfin nous remarquons une particularité dans les données de 1987. Les concentrations en *Dinophysis* pour les jours 72 et 73 sont respectivement de 600 et 46600 cell.l⁻¹. Si l'on fait l'hypothèse qu'en 24 heures le principal facteur de l'augmentation de la concentration est la

reproduction par division cellulaire, alors le taux de croissance de la micro-algue est d'environ 6 j^{-1} . Or les taux de croissance généralement constatés pour plusieurs espèces sont compris entre 0,3 et 2 j^{-1} [25, 26]. Il est raisonnable de penser que des processus hydrodynamiques (advection, accumulation, *etc.*) peuvent être responsable de telles variations. Les phénomènes de microdistribution sont sans doute également à prendre en compte. De plus, il a été montré que les concentrations en *Dinophysis* sont beaucoup plus importantes dans la thermocline. Cette dernière est dépendante de la température et de la salinité, qui sont variables dans le temps, et donc la position de la thermocline varie elle même dans le temps. Enfin le *Dinophysis* est muni de flagelles qui lui permettent de se déplacer. Ces déplacements semblent principalement verticaux. Le *Dinophysis* varie donc dans le temps et dans les trois dimensions de l'espace, et ce à plusieurs échelles temporelles et spatiales. Cependant, on ne dispose d'aucune étude permettant d'apprécier l'ensemble de ces variabilités.

Il nous apparaît que les échecs répétés des différentes méthodes utilisées pour prévoir et comprendre l'évolution des concentrations en *Dinophysis* sont dus à des problèmes d'échantillonnage, pour une part importante. Et pour pouvoir envisager des stratégies d'échantillonnage ayant certaines caractéristiques d'optimalité il est nécessaire de disposer de données fondamentales. Il faut donc penser à réaliser une étude spatiale et temporelle, comportant une surface importante, une maille d'échantillonnage serrée et une fréquence de prélèvement élevée en mesurant tous les paramètres susceptibles d'expliquer la dynamique des concentrations en *Dinophysis*.

5 Conclusion

Notre objectif était de mettre en place un outil mathématique permettant de déclencher de façon optimale l'alerte phytoplanctonique concernant la micro-algue toxique *Dinophysis cf acuminata*. Pour cela nous avons décidé d'utiliser des modèles dynamiques de prévision bayésienne (modèles polynomiaux du premier, second et troisième ordre et modèle de régression multivariée) qui permettent la surveillance en temps réel d'un processus temporel.

Ces modèles se sont avérés insatisfaisants. Il nous semble cependant que les problèmes de modélisations du phénomène sont dus à un manque de connaissances fondamentales et à des stratégies d'échantillonnage inadaptées. C'est pourquoi nous pensons qu'il faudrait effectuer une étude prenant en compte les variabilités spatiales et temporelles des concentrations en *Dinophysis cf acuminata*.

6 Bibliographie

- 1 Yasumoto T., Oshima Y., Sugawara W., Fukuyo Y., Oguri H., Igarashi T., Kujita N., 1980. Identification of *Dinophysis fortii* as the causative organism of Diarrhetic Shellfish Poisoning. Bull. jap. Soc. scient. Fish. 46 : 1405-1411.
- 2 Lassus P., Martin A.G., Maggi P., Berthomé J.P., Langlade A., Bachere E., 1985. Extension du dinoflagellé *Dinophysis cf. acuminata* en Bretagne sud et conséquences pour les cultures marines. Rev. Trav. Inst. Pêches marit. 47 : 122-133.
- 3 Murata M., Shimatani M., Sugitani H., Oshima Y., Yasumoto T., 1982. Isolation and structural elucidation of the causative toxin of the Diarrhetic Shellfish Poisoning. Bull. jap. Soc. scient. Fish. 48 : 549-552.
- 4 Kumagai M., Yanagi T., Murata M., Yasumoto T., Kat M., Lassus P., Rodriguez-Vazquez J.A., 1986. Okadaic acid as the causative toxin of Diarrhetic Shellfish Poisoning in Europe. Agric. Biol. Chem. 50 : 2853-2857.
- 5 Lee J.S., Igarashi T., Fraga S., Dahl E., Hovgaard P., Yasumoto T., 1989. Determination of diarrhetic shellfish toxin in various dinoflagellate species. J. appl. Phycol. 1 : 147-152.
- 6 Sournia A., Belin C., Berland B., Erad Le Denn E., Gentien P., Grzebyk D., Marcaillou Le Baut C., Lassus P., Partensky F., 1991. Le phytoplancton nuisible des côtes de France : de la biologie à la prévention, IFREMER-CNRS : 154 pp.
- 7 Hansen P.J., 1991. *Dinophysis* - a planktonic dinoflagellate genus which can act both as a prey and a predator of a ciliate. Mar. Ecol. Prog. Ser. 69 : 201-204.
- 8 MacKenzie L., 1992. Does *Dinophysis* (Dinophyceae) have a sexual life cycle? J. Phycol. 28 : 399-406.
- 9 Delmas D., Herbland A., Maestrini S.Y., 1992. Environmental conditions which lead to increase in cell density of the toxic dinoflagellates *Dinophysis spp.* in nutrient-rich and nutrient-poor waters of the French Atlantic coast. Mar. Ecol. Prog. Ser. 89 : 53-61.

- 10 Burkholder J.M., Noga E.J., Hobbs C.H., Glasgow H.B., 1992. New "phantom" dinoflagellate is the causative agent of major estuarine fish kills. *Nature* 358 : 407-410.
- 11 Belin C., 1993. Distribution of *Dinophysis spp.* and *Alexandrium minutum* along the French coast since 1984 and their DSP and PSP toxicity levels. In *Toxic Phytoplankton Blooms in the Sea*, Smayda T.J. and Shimizu Y. (eds.) : pp. 469-474, Elsevier Science publishers.
- 12 Lassus P., Bardouil M., Berthomé J.P., Maggi P., Truquet P., Le Déan L., 1988. Seasonal occurrence of *Dinophysis spp.* along the French coast between 1983 and 1987. *Aquat. Living Resour.* 1 : 155-164.
- 13 Delmas D., Herbland A., Maestrini S.Y., 1993. Do *Dinophysis spp.* come from the "open sea" along the French atlantic coast? In *Toxic Phytoplankton Blooms in the Sea*, Smayda T.J. and Shimizu Y. (eds.) : pp. 489-494, Elsevier Science publishers.
- 14 Mer G., 1986. Synthèse des connaissances sur les facteurs pouvant influencer le développement des Blooms de dinoflagellés toxiques - Traitement mathématique des données hydrologiques acquises en baie de Vilaine en 1984 et 1985. Rapport IFREMER-DERO-86.08-MR : 72 pp.
- 15 Cazelles B., Carrat F., Chau N.P., Mary J.M., 1991. Analyses et prédiction de la dynamique de série phytoplanktoniques : études de faisabilité Contrat IFREMER : 38 pp.
- 16 Menesguen A., Lassus P., De Cremoux F., Boutibonnes L., 1990. Modelling *Dinophysis* blooms : a first approach. In *Toxic Marine Phytoplankton*, Granéli E., Sundström B., Edler L. and Anderson D.M (eds) : pp. 195-200, Elsevier Science publishers.
- 17 Harrison P.J., Stevens C.F., 1971. A bayesian approach to short term forecasting. *Oper. Res. Quart.* 22 : 341-362.
- 18 Harrison P.J., Stevens C.F., 1976. Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. (ser. B)* 38 : 205-247.

- 19 Bolstad W.M., 1986. Harrison-Stevens forecasting and the multiprocess Dynamic linear model, *Am. Statistician* 40 : 129-135.
- 20 West M., Harrison P.J., 1989. Bayesian forecasting and dynamic models, Springer series in statistics, Springer-Verlag (ed.) : pp. 703.
- 21 Catherine M., Beliaeff B., Pezeron A., 1991. Etude de salubrité du gisement naturel de coques (*Cerastoderma edule*) de la plage Benoît en baie du Pouliguen - Loire-Atlantique (1989-1990). Rapport IFREMER R.INT.DEL/91.04-Nantes : 85 pp.
- 22 Bulmer M.G., 1974. On fitting the poisson lognormal distribution to species-abundance data. *Biometrics* 30 : 111-124.
- 23 El-Shaarawi A.H., Esterby S. R., Dukta B. J., 1981, Bacterial density in water determined by poisson or negative binomial distribution. *Appl. Environ. Microbiol.* 41 : 107-116.
- 24 Bolstad W.M., 1988. The multiprocess dynamic linear model with biased perturbations : a real time model for growth hormone level. *Biometrika* 75 : 685-692.
- 25 Weiler C.S., Chisholm S. W., 1976. Phased cell division in natural populations of marine dinoflagellates from shipboard cultures. *J. exp. mar. biol. Ecol.* 25 : 239-247.
- 26 Morlaix M., 1992. Croissance et toxicité comparées de deux dinoflagellés : *Alexandrium Tamarens* et *Prorocentrum Lima*. Thèse de doctorat en sciences de l'université de Paris XI Orsay.

TABLE DES ANNEXES

Annexe I Données utilisées	22
I.1 Antifer 1987.....	22
I.2 Antifer 1988.....	22
I.3 Données de la plage Benoît.....	26
Annexe II Valeurs initiales des modèles polynomiaux.....	27
II.1 Modèle du premier ordre	27
II.2 Modèle du second ordre.....	27
II.3 Modèle du troisième ordre.....	27
II.4 Modèle du premier ordre avec intervention	27
II.5 Modèle de regression multivarié	28
Annexe III Localisation du point d'échantillonnage d'Antifer.....	29
Annexe IV Présentation des programmes informatiques.....	30

Annexe I Données utilisées

I.1 Antifer 1987.

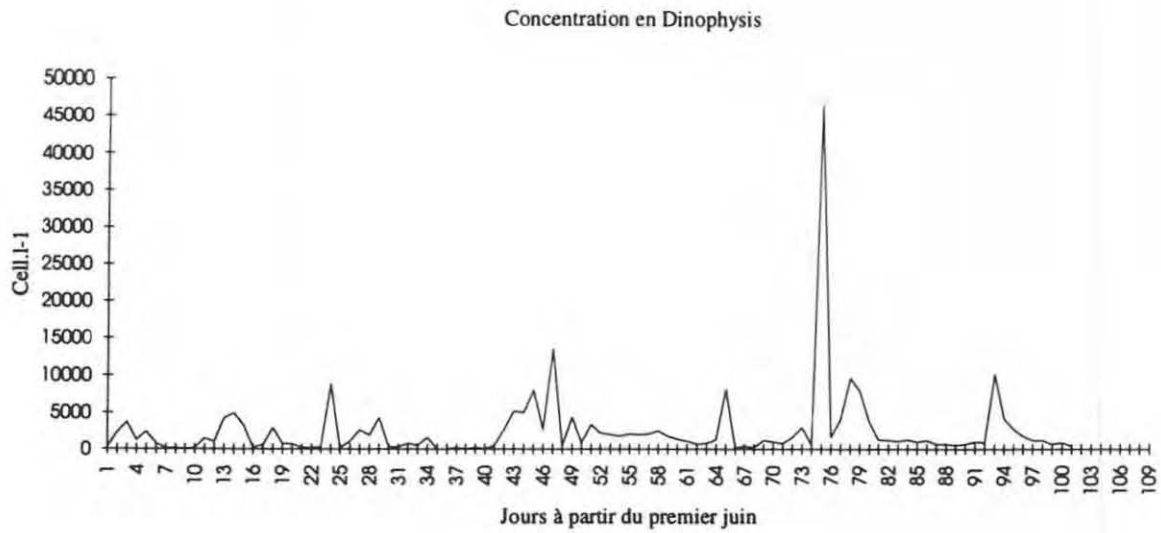


Figure 6 : Concentrations en *Dinophysis cf acuminata* à Antifer (1987).

I.2 Antifer 1988

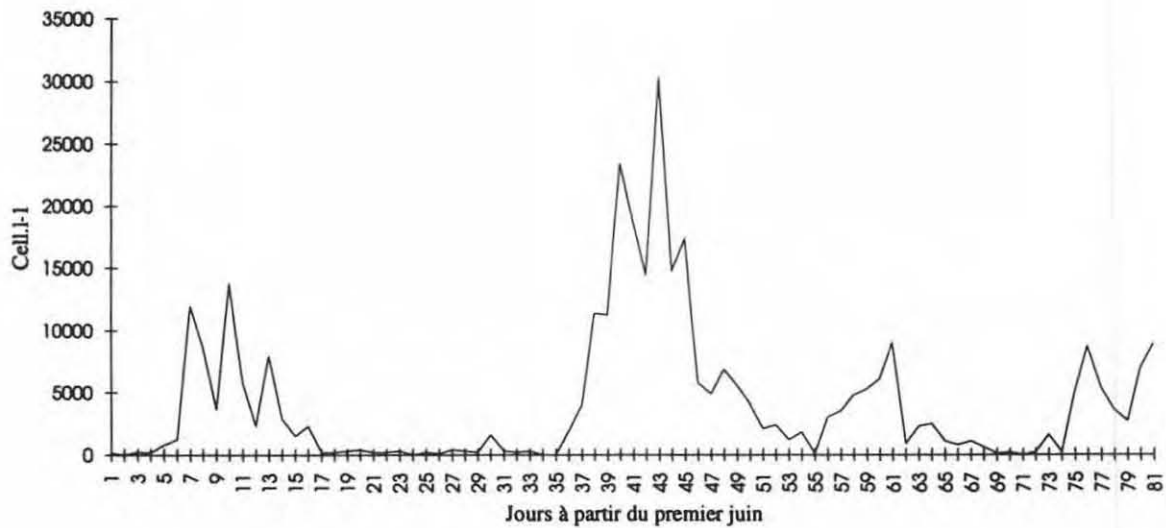


Figure 7 : Concentrations en *Dinophysis cf acuminata* à Antifer (1988).

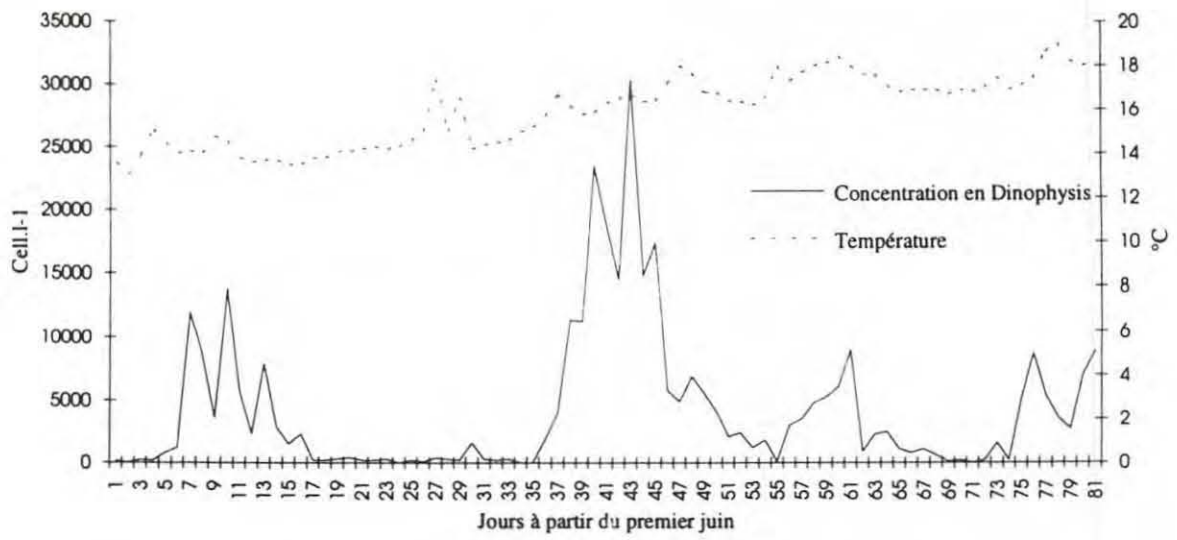


Figure 8 : Concentrations en *Dinophysis cf acuminata* et température à Antifer (1988).

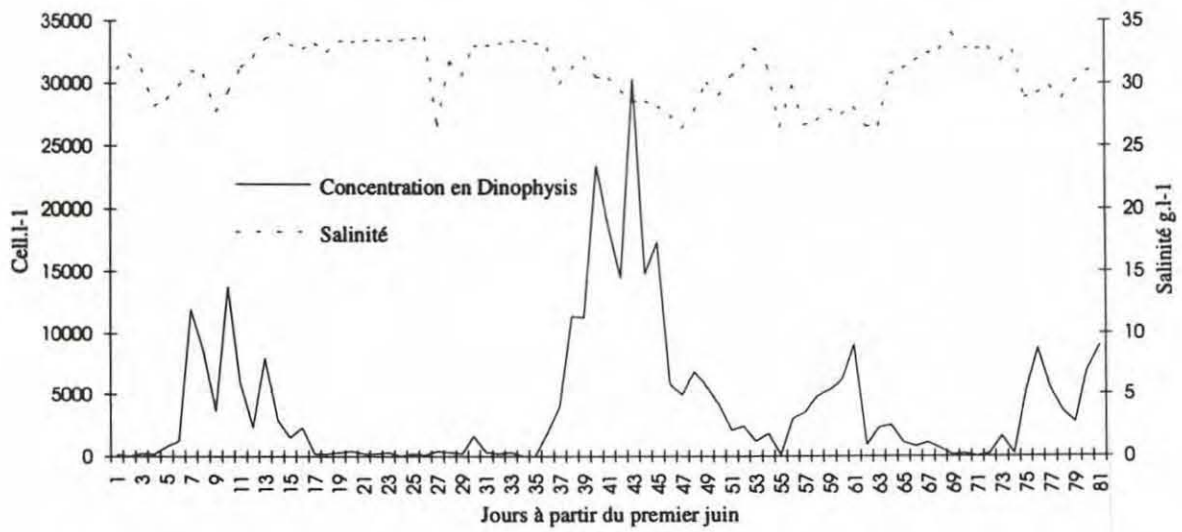


Figure 9 : Concentrations en *Dinophysis cf acuminata* et salinité à Antifer (1988).

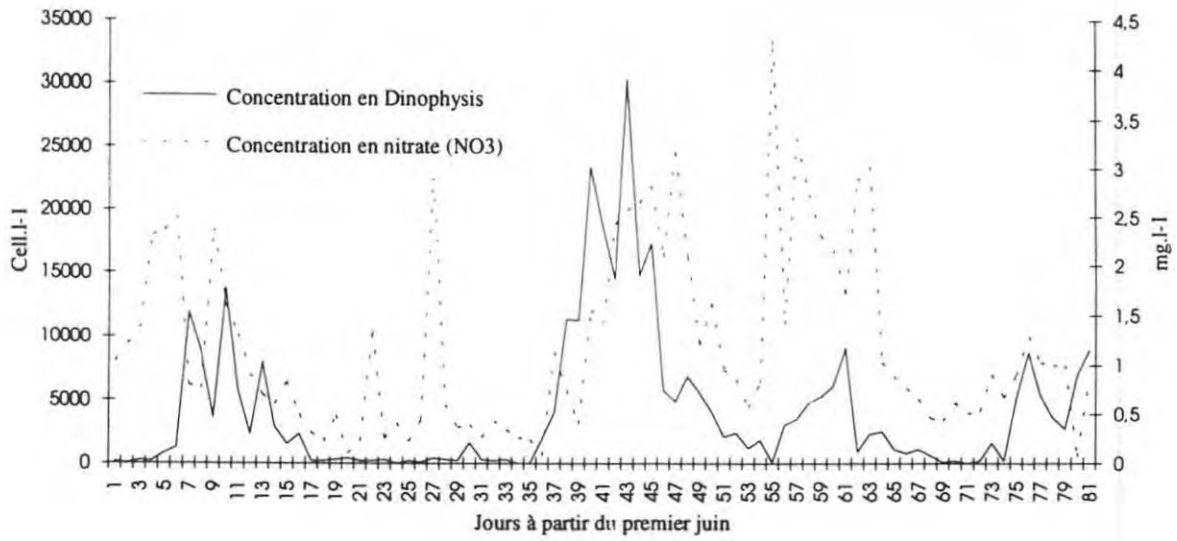


Figure 10 : Concentrations en *Dinophysis cf acuminata* et en nitrate à Antifer (1988).

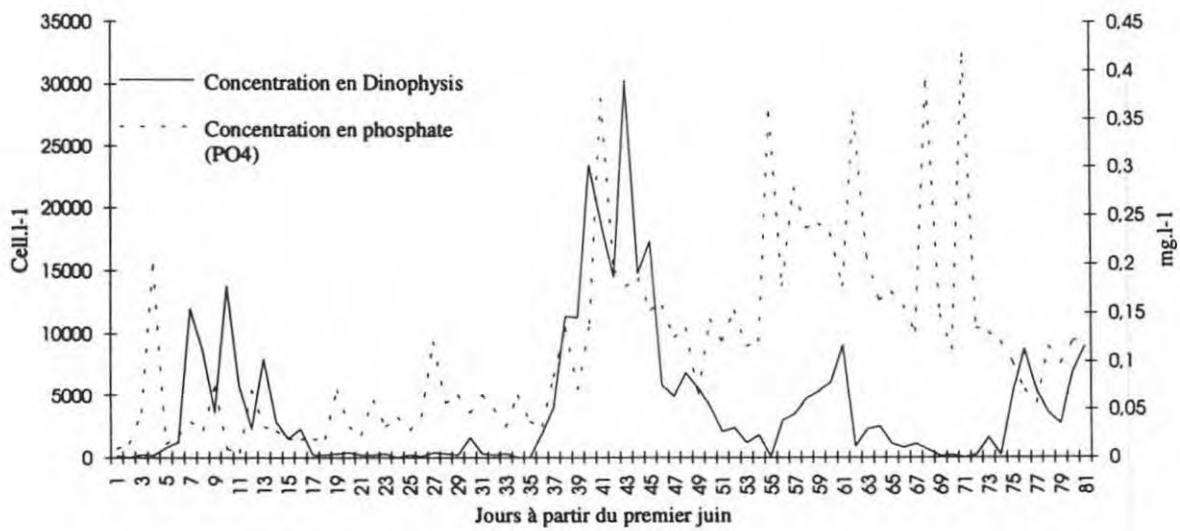


Figure 11 : Concentrations en *Dinophysis cf acuminata* et en phosphate à Antifer (1988).

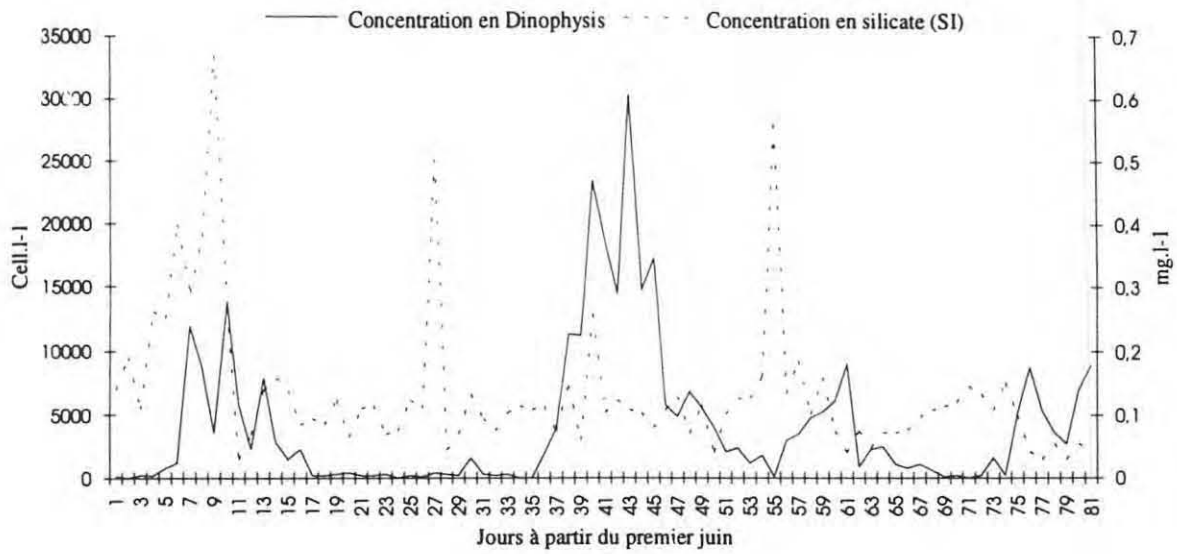


Figure 12 : Concentrations en *Dinophysis cf. acuminata* et en silicate à Antifer (1988).

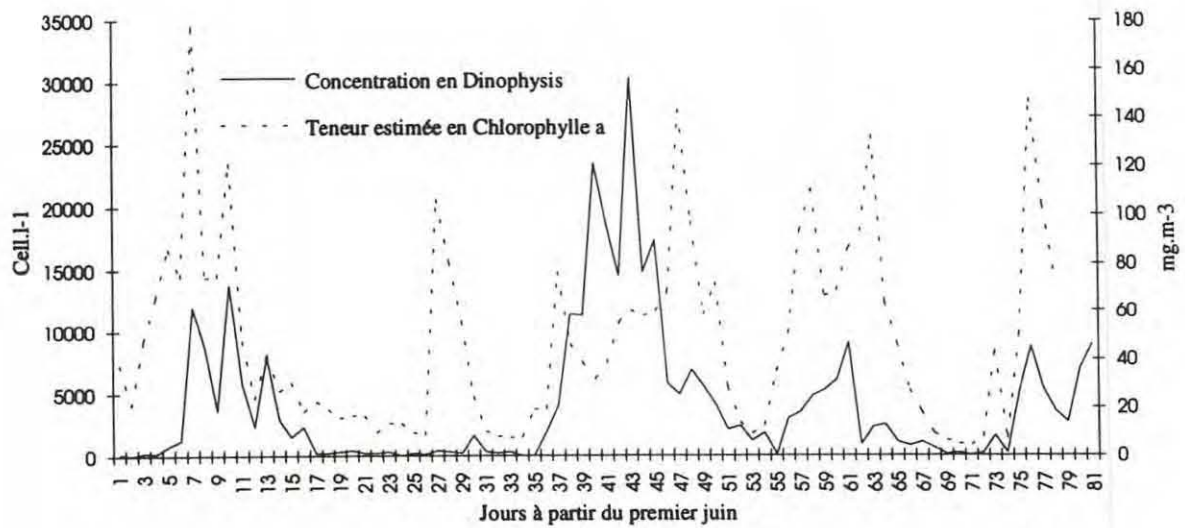


Figure 13 : Concentrations en *Dinophysis cf. acuminata* et teneur estimée en Chlorophylle a à Antifer (1988).

I.3 Données de la plage Benoît

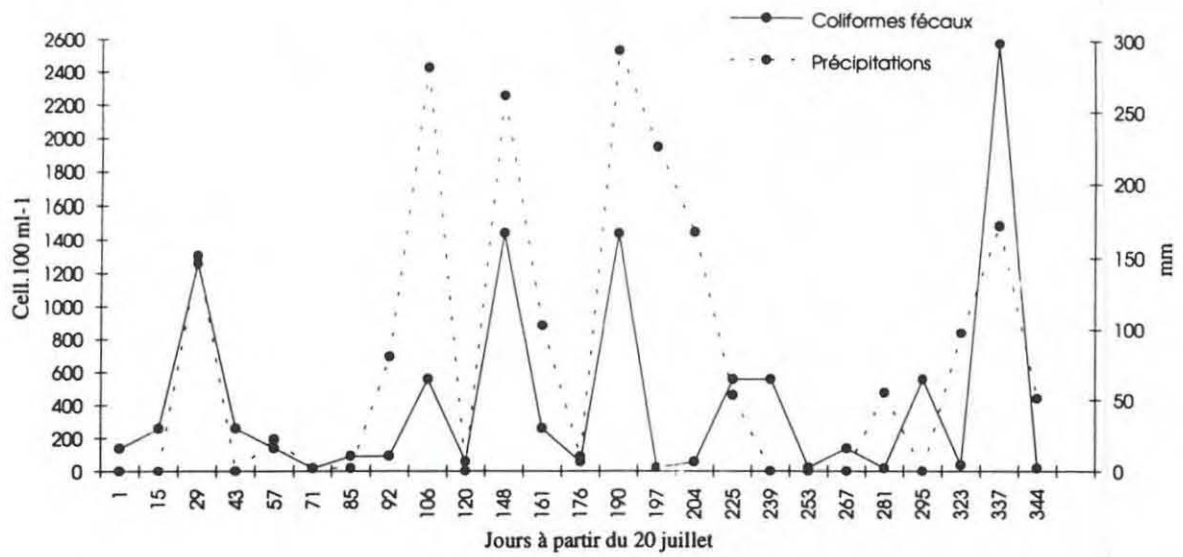


Figure 14 : Taux de coliformes fécaux et précipitations cumulées sur les quatre jours précédant le prélèvement à la plage Benoît (1989).

Annexe II Valeurs initiales des modèles polynomiaux

II.1 Modèle du premier ordre

$$m_0 = 2500$$

$$C_0 = W = 1057600$$

$$V = 330500$$

II.2 Modèle du second ordre

$$\underline{m}_0 = \begin{pmatrix} 2500 \\ 0 \end{pmatrix}$$

$$\underline{C}_0 = \underline{W}$$

$$\underline{\omega} = \begin{pmatrix} 62500 \\ 390625 \end{pmatrix}$$

$$\underline{W} = \begin{pmatrix} 453125 & 390625 \\ 390625 & 390625 \end{pmatrix}$$

$$V = 312500$$

II.3 Modèle du troisième ordre

$$\underline{m}_0 = \begin{pmatrix} 2500 \\ 0 \\ 0 \end{pmatrix}$$

$$\underline{C}_0 = \underline{W}$$

$$\underline{\omega} = \begin{pmatrix} 78125 \\ 28750 \\ 15625 \end{pmatrix}$$

$$\underline{W} = \begin{pmatrix} 765625 & 687500 & 15625 \\ 687500 & 687500 & 15625 \\ 15625 & 15625 & 15625 \end{pmatrix}$$

$$V = 312500$$

II.4 Modèle du premier ordre avec intervention

$$m_0 = 130$$

$$C_0 = W = 95698$$

$$V = 191396$$

II.5 Modèle de regression multivarié

$$\underline{m}_0 = \begin{pmatrix} 1774 \\ 3366 \\ 1717 \end{pmatrix}$$

$$\underline{C}_0 = \underline{W}$$

$$\underline{W} = \begin{pmatrix} 1132096 & -3182668 & -343320 \\ -3182668 & 50470000 & -2280245 \\ -343320 & -2280245 & 515524 \end{pmatrix}$$

$$V = 1562600$$

Annexe III Localisation du point d'échantillonnage d'Antifer

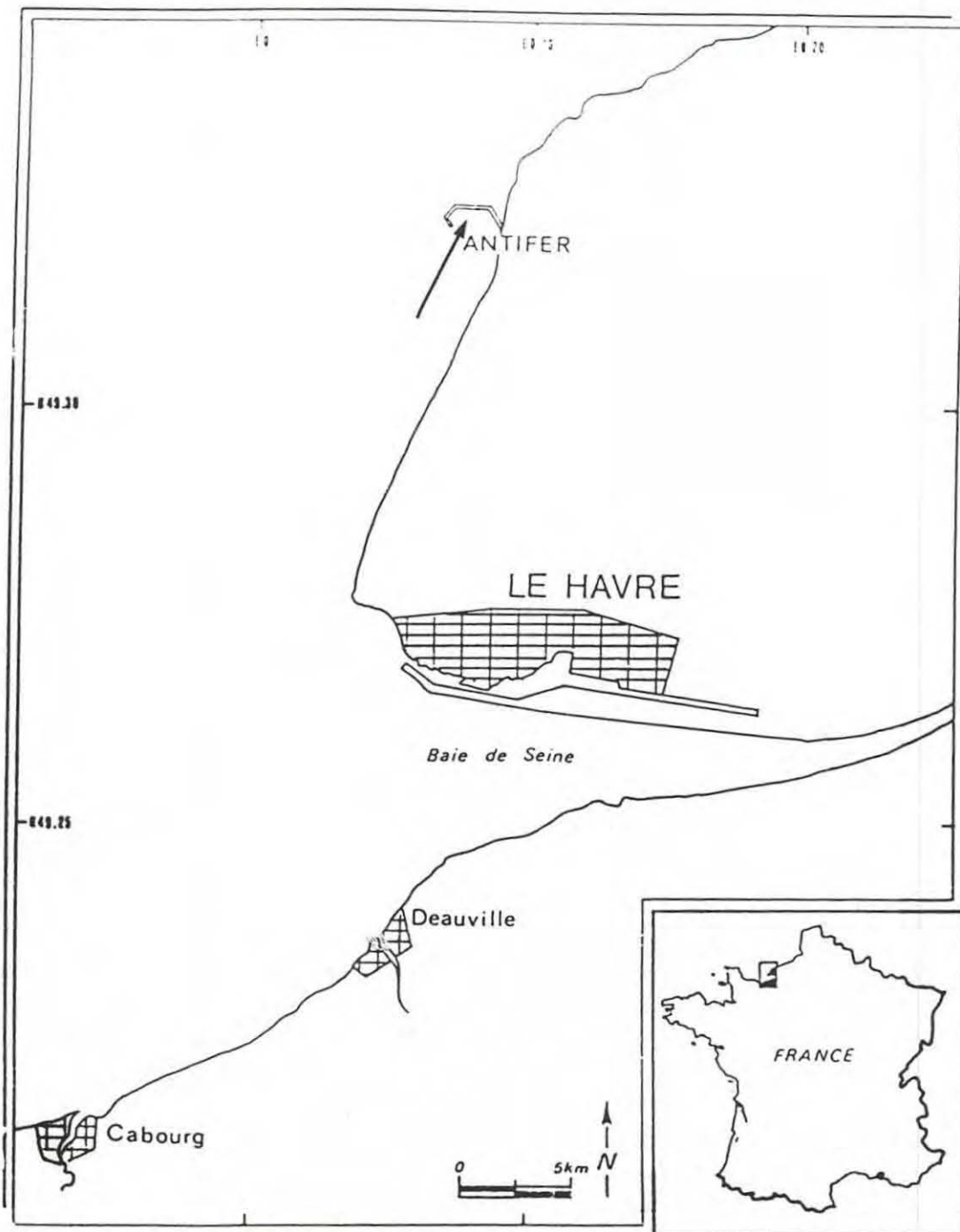


Figure 15 : Situation géographique du terminal d'Antifer (Baie de Seine) - D'après : De Cremoux, 1988. Recherche des facteurs conditionnant le développement de *Dinophysis* par simulations.

Annexe IV Présentation des programmes

Les programmes ont été réalisés en langage C sous UNIX sur une station SUN. Ils sont disponibles auprès de Benoît Beliaeff (Centre de Nantes : DEL/QM).

Les programmes des modèles polynomiaux s'appellent respectivement `FirstOrderPolynomial`, `SecondOrderPolynomial` et `ThirdOrderpolynomial`. Le programme de régression multilinéaire s'appelle `DlmMultireg`. Ces programmes prennent en entrée des fichiers textes comportant des variables en colonne et des observations en ligne. En sortie, ils fournissent les éléments principaux des calculs. L'utilisateur spécifie les paramètres initiaux et les variances d'évolution et d'observation. Par ailleurs, il existe un programme graphique utilisant la librairie `xwindows`, `graph`, qui permet de visualiser les courbes de variables d'un fichier texte.