Méthodes mathématiques d'analyse et de modélisation appliquées à l'environnement.

Dr. Ir. Éric J.M. DELHEZ

Septembre 2008

Chapitre 1

Concepts et outils de l'analyse mathématique.

1.1 Fonction et relation.

La description des systèmes passe généralement par l'association de grandeurs entre elles, ce qui se traduit mathématiquement par la définition de *fonctions* et de *relations*.

Une fonction f est une loi qui, à tout élément x d'un ensemble $\mathtt{E}$, appelé *domaine de définition* de la fonction, associe un et un seul élément $f(x)$ d'un ensemble $\mathtt{F}$. Les ensembles $\mathtt{E}$ et $\mathtt{F}$ peuvent contenir des éléments très différents de par leur nature (individus, nombre, tenseur,...) ou leur interprétation (temps, température, vitesse, espèce,...). L'élément important dans la définition d'une fonctions, c'est l'association d'un élément unique de $\mathtt{F}$ à tout élément de $\mathtt{E}$. Cette association peut parfois être explicitée au moyen d'une formule mathématique, d'une table, d'un graphe, ... Ce n'est cependant pas nécessairement le cas.

La notion de relation généralise celle de fonction. Dans une relation, les éléments du domaine de définition $\mathtt{E}$ peuvent être mis en correspondance avec plusieurs éléments de $\mathtt{F}$. Certains éléments du domaine de définition peuvent également n'être associé à aucun élément de $\mathtt{F}$. Par exemple, on peut définir une relation entre un ensemble de sites $\mathtt{E}$ et un ensemble $\mathtt{F}$ d'espèces indicatrices en associant à chaque site les espèces qui y sont présentes.

Bien que souvent généralisables à des espaces plus généraux, la plupart des outils de l'analyse mathématique sont particulièrement bien adaptés à l'étude des fonctions réelles d'une ou plusieurs variables réelles, *i.e.* $\mathtt{E}, \mathtt{F} \subset \mathbb{R}^n$. On s'efforcera donc autant que possible de traduire sous cette forme les propriétés physiques, biologiques ou chimiques des systèmes étudiés, donnant ainsi accès à une description quantitative de l'état des systèmes et des processus qui s'y déroulent.

1.2 Limite et comportement asymptotique.

Le calcul de la *limite* d'une fonction constitue l'outil fondamental de l'analyse mathématique continue. Mathématiquement, on écrit :

$$\begin{gathered}\lim_{x\to x_0} f(x) = a \in \mathbb{R}^n \\ \Updownarrow \\ (\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in E, 0 < |x - x_0| \leq \delta) : |f(x) - a| \leq \varepsilon\end{gathered} \tag{1.1}$$

L'existence d'une limite finie de f pour x tendant vers x_0 signifie que l'on peut rendre les valeurs de la fonction $f(x)$ aussi proches que l'on veut de la constante a en considérant des points $x \in E$ suffisamment proches de x_0 (sauf éventuellement x_0).

Si f est à valeur dans $\mathbb{R}$, on écrira

$$\lim_{x\to x_0} f(x) = -\infty \qquad \text{ou} \qquad \lim_{x\to x_0} f(x) = +\infty \tag{1.2}$$

pour signifier que les valeurs de f sont non bornées au voisinage de x_0.

La définition (1.1) est générale. Dans la suite, nous considérerons quasi exclusivement des fonctions réelles d'une seule variable. Dans ce cas, l'existence de limites particulières se traduit par des *asymptotes* dans le graphe de f.

Ainsi, si

$$\lim_{x\to x_0^+} f(x) = \pm\infty \tag{1.3}$$

en tous les points $x > x_0$ suffisamment proches de x_0, la fonction croît ou décroît indéfiniment et se rapproche aussi près que l'on désire de la droite $x = x_0$. De même, si

$$\lim_{x\to x_0^-} f(x) = \pm\infty \tag{1.4}$$

la fonction $f(x)$ approche la droite $x = x_0$ pour des valeurs de x inférieures à x_0. On dit dans les deux cas que le graphe de f comporte une *asymptote verticale* en $x = x_0$. Remarquons que ce comportement peut être différent de part et d'autre de x_0.

De même, si

$$\lim_{x\to+\infty} f(x) = a \text{ fini} \tag{1.5}$$

ou si

$$\lim_{x\to-\infty} f(x) = a \text{ fini} \tag{1.6}$$

la fonction se rapproche indéfiniment de la droite horizontale $y = a$ à mesure que x croît ou décroît. On dit que le graphe de la fonction $y = f(x)$ possède une *asymptote horizontale* $y = a$. Dans le cas où les limites peuvent s'écrire

$$\lim_{x\to\pm\infty} f(x) = a^- \qquad \text{ou} \qquad \lim_{x\to\pm\infty} f(x) = a^+ \tag{1.7}$$

f approche l'asymptote horizontale par le dessous ou par le dessus.

EXEMPLE 1.1 Considérons la fonction de Michaelis-Menten décrivant la variation du taux de croissance μ d'une espèce phytoplanctonique en fonction de la concentration en nutriment $[N]$ dans le milieu :

$$\mu([N]) = \mu_{max}\frac{[N]}{[N]+\kappa}$$

où κ désigne la constante de demi-saturation (Fig. 1.1).

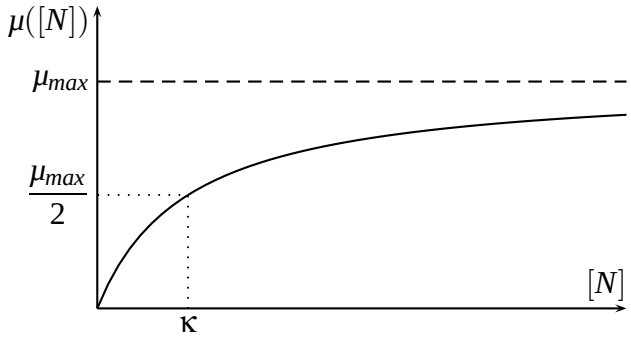


FIG. 1.1

On calcule aisément

$$\lim_{[N]\to 0}\mu([N]) = 0$$

et

$$\lim_{[N]\to +\infty}\mu([N]) = \mu_{max}$$

Le premier résultat indique que le taux de croissance est très faible lorsque la concentration en nutriment est proche de zéro (Plus exactement, le taux de croissance peut être rendu arbitrairement petit en considérant des concentrations proches de zéro.).

Le second résultat montre que le taux de croissance est proche de la valeur μ_{max} lorsque les concentrations sont élevées. Écrivant $\mu([N])$ sous la forme

$$\mu([N]) = \mu_{max}\left(1-\frac{1}{[N]/\kappa+1}\right) < \mu_{max} \qquad (\flat)$$

on vérifie aisément que $\mu([N])$ approche sa valeur limite par valeur inférieure, *i.e.* que le graphe de $\mu([N])$ est situé sous l'asymptote. Ceci s'écrit

$$\lim_{[N]\to +\infty}\mu([N]) = \mu_{max}^{-}$$

L'écriture de la loi de Michaelis-Menten sous la forme $\flat$ montre également que le comportement de μ dépend essentiellement du rapport $[N]/\kappa$. Nous reviendrons sur ce point lors de l'étude des variables adimensionnelles. Nous pouvons cependant déjà remarquer que la constante κ apparaît comme une concentration caractéristique par rapport à laquelle la concentration $[N]$ peut être comparée. Ainsi, l'asymptote horizontale $\mu = \mu_{max}$ est bien approchée lorsque la concentration $[N]$ est grande *par rapport* à κ. De même, le taux de croissance sera faible lorsque $[N]$ est lui-même petit *par rapport* à la constante de demi-saturation. $\diamond$

La notion d'*asymptote oblique* est introduite dans le cas où on peut trouver des réels a et b finis tels que

$$\lim_{x\to+\infty} f(x) = \pm\infty, \qquad \lim_{x\to+\infty} \frac{f(x)}{x} = a \qquad \text{et} \qquad \lim_{x\to+\infty} \Big[f(x) - ax \Big] = b \tag{1.8}$$

ou

$$\lim_{x\to-\infty} f(x) = \pm\infty, \qquad \lim_{x\to-\infty} \frac{f(x)}{x} = a \qquad \text{et} \qquad \lim_{x\to-\infty} \Big[f(x) - ax \Big] = b \tag{1.9}$$

Dans ce cas, la fonction f finit par se rapprocher indéfiniment de la droite oblique $y = ax + b$ qui est l'équation de l'asymptote oblique de f. Dans les cas où

$$\lim_{x\to\pm\infty} \Big[f(x) - ax \Big] = b^- \qquad \text{ou} \qquad \lim_{x\to\pm\infty} \Big[f(x) - ax \Big] = b^+ \tag{1.10}$$

on peut encore préciser si f approche l'asymptote oblique par le dessous ou par le dessus.

EXEMPLE 1.2 Esquissons le graphe de la fonction

$$f(x) = \frac{x^2 - x - 6}{x + 1}$$

Cette fonction est définie pour tout x réel à l'exception de $x = -1$. En ce point, on a

$$\lim_{x\to-1^-} f(x) = +\infty \qquad \text{et} \qquad \lim_{x\to-1^+} f(x) = -\infty$$

Le graphe de f comporte donc une asymptote verticale en $x = -1$.

À l'infini, on obtient

$$\lim_{x\to-\infty} f(x) = -\infty \qquad \text{et} \qquad \lim_{x\to+\infty} f(x) = +\infty$$

Si on calcule les limites

$$\lim_{x\to-\infty} \frac{f(x)}{x} = 1 = \lim_{x\to+\infty} \frac{f(x)}{x}$$

et

$$\lim_{x\to-\infty} (f(x) - x) = -2 = \lim_{x\to+\infty} (f(x) - x)$$

on en déduit la présence d'une asymptote oblique $y = x - 2$ aussi bien en $-\infty$ qu'en $+\infty$. Ce résultat peut aussi être obtenu directement en réécrivant la fonction sous la forme

$$f(x) = \frac{x^2 - x - 6}{x + 1} = x - 2 - \frac{4}{x + 1}$$

où le dernier terme tend vers zéro lorsque x tend vers $\pm\infty$. Dès lors, pour de grandes valeurs de $|x|$, $4/(x+1)$ devient négligeable vis à vis de $x - 2$ et $f(x)$ se rapproche indéfiniment de l'asymptote oblique. Lorsque x est grand et positif, le terme $4/(x+1)$ est petit et positif de sorte que $f(x)$ approche l'asymptote par défaut. Pour x négatif, c'est l'inverse qui se produit et $f(x)$ est alors légèrement supérieure à $x - 2$.

Si on ajoute à ces résultats que $f(x)$ s'annule en $x = -2$ et $x = 3$, une esquisse du graphique de $f(x)$ est donné par

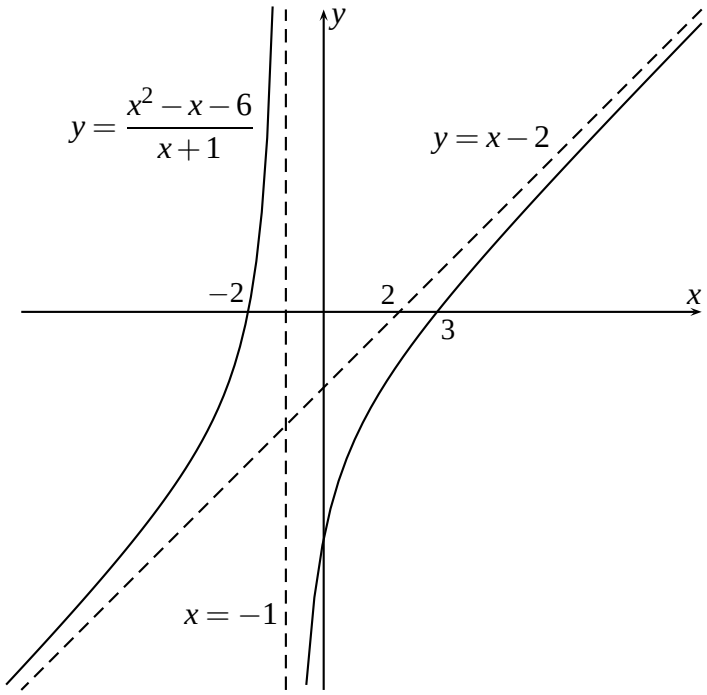


FIG. 1.2

◇

Lorsqu'une fonction possède une asymptote en $x \to +\infty$, l'écart entre le graphique de la fonction et celui de son asymptote peut être rendu arbitrairement petit à condition de prendre x suffisamment grand. Plus généralement, on peut comparer le comportement d'une fonction f à celui d'une autre fonction g (dont la forme analytique est en général plus simple ou mieux connue que celle de f). Ainsi, une fonction f est dite *asymptotique* ou *équivalente* à g au voisinage de x_0, ce que l'on note $f(x) \sim g(x)$, $(x \to x_0)$ lorsque

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1 \tag{1.11}$$

La relation d'équivalence $\sim$ est symétrique :

$$f(x) \sim g(x) \qquad \Leftrightarrow \qquad g(x) \sim f(x) \tag{1.12}$$

Si une fonction f possède une limite finie $a \neq 0$ en un point x_0 de son domaine de définition, elle est asymptotique à la fonction constante $g(x) = a$. De même, si elle

présente une asymptote oblique $y = ax + b$ ou une asymptote horizontale $y = a$, avec $a \neq 0$, elle est asymptotique à cette asymptote.

La notion de fonction asymptotique permet cependant d'aller au-delà des notions de limite et d'asymptote en précisant, par exemple, la façon dont les valeurs d'une fonction tendent vers zéro ou vers l'infini.

EXEMPLE 1.3 Reprenons l'exemple de la fonction de limitation Michaelis-Menten

$$\mu([N]) = \mu_{max}\frac{[N]}{[N]+\kappa}$$

Nous avons vu que

$$\lim_{[N]\to\infty} \mu([N]) = \mu_{max}$$

De façon alternative, on peut également décrire le comportement de μ pour les grandes concentrations de nutriments en disant que $\mu([N])$ est asymptotique à μ_{max} pour $[N] \to +\infty$ (ou, mieux encore, pour $[N]/\kappa \to +\infty$) :

$$\mu([N]) \sim \mu_{max}, \qquad [N] \to +\infty$$

D'autre part, le résultat

$$\lim_{[N]\to 0} \mu([N]) = 0$$

est relativement peu précis au sujet du comportement de μ pour les faibles concentrations : le taux de croissance peut en effet tendre vers zéro plus ou moins rapidement. Il est beaucoup plus instructif de remarquer que

$$\lim_{[N]\to 0} \frac{\mu([N])}{[N]} = \lim_{[N]\to 0} \mu_{max}\frac{1}{\kappa+[N]} = \frac{\mu_{max}}{\kappa}$$

ce qui permet d'écrire

$$\mu([N]) \sim \mu_{max}\frac{[N]}{\kappa}, \qquad [N] \to 0$$

Cette dernière expression permet en effet de remplacer la dépendance réelle de μ en $[N]$ par une fonction linéaire dans un voisinage de zéro. ◇

EXEMPLE 1.4 La fonction de l'exemple 1.2,

$$f(x) = \frac{x^2 - x - 6}{x+1}$$

possède des limites infinies à gauche et à droite de $x = -1$. La notion de fonction asymptotique permet de préciser ce comportement. Isolant la partie singulière de f au voisinage de $x = -1$, on a

$$f(x) \sim -\frac{4}{x+1}, \quad (x \to -1)$$

En effet,

$$\lim_{x\to -1} \frac{f(x)}{\dfrac{(-4)}{x+1}} = \lim_{x\to -1} \frac{x^2-x-6}{(-4)} = 1$$

De même, le comportement à l'infini peut être décrit par

$$f(x) \sim x-2, \qquad x \to \pm\infty$$

◇

EXEMPLE 1.5 La vitesse de propagation (vitesse de phase) des ondes de longueur d'onde L dans un basin de profondeur D est donnée par[1]

$$c = \sqrt{\frac{gL}{2\pi}\,\mathrm{th}\,\frac{2\pi D}{L}}$$

où g désigne l'accélération de la pesanteur. Si la longueur d'onde L est fixée, c est donc une fonction croissante de la profondeur.

Le calcul de la limite

$$\lim_{D\to\infty} \sqrt{\frac{gL}{2\pi}\,\mathrm{th}\,\frac{2\pi D}{L}} = \sqrt{\frac{gL}{2\pi}}$$

montre que

$$c \sim \sqrt{\frac{gL}{2\pi}}, \qquad D \to \infty$$

i.e. la vitesse des ondes en eaux profondes (les eaux peuvent être qualifiées de profondes si $2\pi D/L \gg 1$) est indépendante de la profondeur.

Pour des faibles profondeurs, le calcul de la limite

$$\lim_{D\to 0} \sqrt{\frac{gL}{2\pi}\,\mathrm{th}\,\frac{2\pi D}{L}} = 0$$

est de peu d'utilité : les ondes ne peuvent se propager si la profondeur est nulle. Tenant compte de[2]

$$\mathrm{th}\,x \sim x, \qquad x \to 0$$

on obtient

$$c = \sqrt{\frac{gL}{2\pi}\,\mathrm{th}\,\frac{2\pi D}{L}} \sim \sqrt{\frac{gL}{2\pi}\,\frac{2\pi D}{L}} = \sqrt{gD}, \qquad D \to 0$$

[1] Rappelons les définitions des fonctions hyperboliques

$$\mathrm{sh}\,x = \frac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}, \qquad \mathrm{ch}\,x = \frac{\mathrm{e}^x + \mathrm{e}^{-x}}{2}, \qquad \mathrm{th}\,x = \frac{\mathrm{sh}\,x}{\mathrm{ch}\,x} \tag{1.13}$$

[2] On peut vérifier cette propriété en utilisant le théorème de l'Hospital pour lever l'indétermination de la limite

$$\lim_{x\to 0} \frac{\mathrm{th}\,x}{x} = \left(\frac{0}{0}\right) = \lim_{x\to 0} \frac{(\mathrm{th}\,x)'}{(x)'} = \lim_{x\to 0} \frac{1+\mathrm{th}^2 x}{1} = 1$$

.

Là où la profondeur est faible *par rapport à la longueur d'onde*, la vitesse des ondes est proportionnelle à la racine carrée de la profondeur. Cette dépendance peut être observée sur une carte de propagation de la marée : la distance entre les lignes cotidales diminue avec la profondeur. Elle est aussi responsable de l'alignement des fronts de vagues avec les isobathes à l'approche de la côte.

En pratique, on considère généralement que les eaux profondes et peu profondes correspondent respectivement à $D > L/2$ et $D < L/20$.

◇

Si on peut remplacer une expression compliquée f par une fonction simple g qui lui est asymptotique, c'est parce que l'on considère que la différence entre ces deux fonctions f et g est négligeable. Ce concept de différence négligeable peut être défini de façon rigoureuse. Une fonction f est dite *négligeable* par rapport à une fonction g au voisinage de x_0 (qui peut être infini), ce que l'on note $f(x) = o(g(x)),\ (x \to x_0)$, lorsque

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0 \tag{1.14}$$

Avec cette définition,

$$f(x) \sim g(x) \qquad \text{si et seulement si} \qquad f(x) - g(x) = o(g(x)), \quad (x \to x_0) \tag{1.15}$$

EXEMPLE 1.6 Les puissances de x définissent une *échelle de mesure* telle que

$$\forall n \in \mathbb{Z},\ \forall k \in \mathbb{N}_0,\ x^n = o(x^{n+k}), (x \to +\infty)$$

En effet,

$$\lim_{x \to +\infty} \frac{x^n}{x^{n+k}} = \lim_{x \to +\infty} \frac{1}{x^k} = 0$$

Inversement,

$$\forall n \in \mathbb{Z},\ \forall k \in \mathbb{N}_0,\ x^{n+k} = o(x^n), \quad (x \to 0)$$

En effet,

$$\lim_{x \to 0} \frac{x^{n+k}}{x^n} = \lim_{x \to 0} x^k = 0$$

◇

Enfin, on dit qu'une fonction f est *au plus de l'ordre de g* dans un voisinage de x_0, ce que l'on note $f(x) = O[g(x)],\ (x \to x_0)$, lorsque la limite de f/g existe et est finie :

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = M \text{ fini} \qquad \Rightarrow \qquad f(x) = O[g(x)],\ (x \to x_0) \tag{1.16}$$

En particulier, $f = O(1), (x \to x_0)$ signifie que f est bornée au voisinage de x_0.

L'expression $f(x) = O[g(x)]$ est une affirmation moins forte que $f(x) \sim g(x)$ ou $f(x) = o(g(x))$. En effet, $f(x) \sim g(x)$ ou $f(x) = o(g(x))$ implique $f(x) = O[g(x)]$. La réciproque est fausse.

Les notations o et O sont généralement utilisées pour exprimer l'ordre de grandeur des termes négligés dans un développement mathématique en précisant de la sorte le comportement relatif de deux fonctions.

Les relations o, O et $\sim$ admettent les règles simples de manipulation :

$$f_1 = O(g), f_2 = O(g) \quad \Rightarrow \quad \alpha f_1 + \beta f_2 = O(g) \ \ \forall \alpha, \beta \in \mathbb{C} \tag{1.17}$$

$$f_1 = O(g_1), f_2 = O(g_2) \quad \Rightarrow \quad f_1 f_2 = O(g_1 g_2) \tag{1.18}$$

$$f_1 = o(g), f_2 = o(g) \quad \Rightarrow \quad \alpha f_1 + \beta f_2 = o(g) \ \ \forall \alpha, \beta \in \mathbb{C} \tag{1.19}$$

$$f_1 = o(g_1), f_2 = o(g_2) \quad \Rightarrow \quad f_1 f_2 = o(g_1 g_2) \tag{1.20}$$

$$\left.\begin{array}{c} f = o(g), g = O(h) \\ \text{ou} \\ f = O(g), g = o(h) \end{array}\right\} \quad \Rightarrow \quad f = o(h) \tag{1.21}$$

$$f_1 \sim g_1, g_1 \sim g_2 \quad \Rightarrow \quad f_1 \sim g_2 \tag{1.22}$$

$$f_1 \sim g_1, f_2 \sim g_2 \quad \Rightarrow \quad f_1 f_2 \sim g_1 g_2 \tag{1.23}$$

$$f \sim g, 1/f \text{ définie en } x_0 \quad \Rightarrow \quad 1/f \sim 1/g \tag{1.24}$$

1.3 Dérivée.

La *dérivée* de la fonction d'une variable réelle f au point $x^\star$ (appartenant au domaine de définition de f) est donnée par

$$f'(x^\star) = \lim_{\Delta x \to 0} \frac{f(x^\star + \Delta x) - f(x^\star)}{\Delta x} \tag{1.25}$$

si cette limite existe et est finie. On peut aussi écrire

$$f'(x^\star) = \frac{df}{dx}(x^\star) = (Df)(x^\star) = \lim_{x \to x^\star} \frac{f(x) - f(x^\star)}{x - x^\star} \tag{1.26}$$

D'un point de vue géométrique, la dérivée peut être interprétée comme la pente de la tangente au graphe de la fonction f. En effet, si P désigne le point du graphe de f

(Fig. 1.3) correspondant au point $(x^\star, f(x^\star))$ et si on considère un point Q voisin de P, les coordonnées de celui-ci sont $(x^\star+\Delta x, f(x^\star+\Delta x))$ de sorte que l'*accroissement* Δx donné à x produit un accroissement

$$\Delta f = f(x^\star+\Delta x) - f(x^\star) \tag{1.27}$$

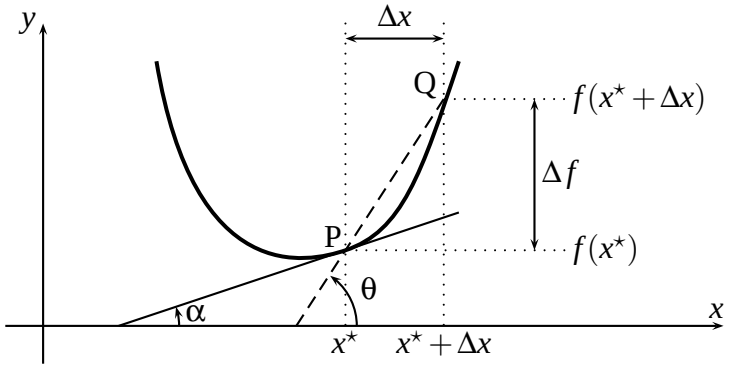


FIG. 1.3

Le *quotient différentiel*

$$\frac{\Delta f}{\Delta x} = \frac{f(x^\star+\Delta x) - f(x^\star)}{\Delta x} = \operatorname{tg}\theta \tag{1.28}$$

représente alors la pente de la droite joignant P et Q ou, si on préfère, le taux moyen de variation de f entre les deux points. Lorsque Q se rapproche indéfiniment de P, la droite PQ tend vers la tangente à la courbe en P et le quotient différentiel (1.28) tend vers la pente $\operatorname{tg}\alpha$ de cette tangente.

Si la fonction f est dérivable et si sa dérivée f' est elle-même dérivable, la dérivée de f' se note

$$f''(x) = \frac{d^2 f}{dx^2}(x) = D^2 f(x) \tag{1.29}$$

et est appelée la *dérivée seconde* de f. D'une façon générale,

$$f^{(n)}(x) = \frac{d^n f}{dx^n}(x) = D^n f(x) \tag{1.30}$$

représente la *dérivée d'ordre n* de f (si cette expression à un sens).

La dérivée d'une fonction nous renseigne sur la sensibilité de la grandeur mesurée par cette fonction aux variations de la variable. Implicitement, cela revient à interpréter la dérivée de f comme le coefficient de proportionnalité qui relie les variations des variables

indépendante x et dépendante f, *i.e.* à remplacer le graphe de f par celui de sa tangente au point d'évaluation de la dérivée. Ceci peut être explicité en écrivant,

$$f(x) = f(x^\star) + f'(x^\star)(x - x^\star) + o(x - x^\star), \qquad x \to x^\star \tag{1.31}$$

Cette relation signifie que l'erreur commise en remplaçant f par l'approximation linéaire correspondant à sa tangente en $x^\star$ décroît plus vite que $x - x^\star$ lorsqu'on se rapproche de $x^\star$. Remplacer f par sa linéarisation au voisinage de $x^\star$ est donc entaché d'une erreur d'autant plus petite que l'on se trouve proche de $x^\star$.

EXEMPLE 1.7 Considérons à nouveau la loi de Michaelis-Menten

$$\mu = \mu_{max} \frac{[N]}{\kappa + [N]}$$

et étudions la sensibilité de μ aux variations de la concentration de nutriments N. On a

$$\frac{d\mu}{d[N]} = \frac{\mu_{max}\kappa}{([N] + \kappa)^2}$$

et donc

$$\mu([N]) \sim \mu([N]^\star) + \frac{\mu_{max}\kappa}{([N]^\star + \kappa)^2} ([N] - [N]^\star) + o([N] - [N]^\star), \qquad ([N] \to [N]^\star)$$

Si $[N]^\star$ est grand, on constate que μ est pratiquement indépendant de $[N]$, en accord avec l'asymptote horizontale identifiée précédemment. Pour $[N]^\star = 0$, on retrouve

$$\mu([N]) = \frac{\mu_{max}}{\kappa} [N] + o([N])$$

◇

EXEMPLE 1.8 Les thermistors (XBT) permettent de mesurer la température de l'eau en se basant sur la variation de la résistance électrique R d'une électrode avec la température T. On a, par exemple,

$$R(T) = R_0 \exp\left[\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

où R_0 désigne la résistance à la température T_0 et β est une constante. Les températures T et T_0 sont exprimées en Kelvin.

La sensibilité de la résistance aux variations de T est donnée par

$$R'(T) = -\frac{R_0\beta}{T^2} \exp\left[\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

où le signe négatif montre que la résistance diminue lorsque la température augmente.

Si on travaille à une température T proche de T_0, on aura

$$R(T) = R(T_0) + R'(T_0)(T - T_0) + o(T - T_0)$$

soit

$$\frac{R(T)}{R_0} = 1 - \frac{\beta}{T_0^2}(T - T_0) + o(T - T_0), \qquad (T \to T_0)$$

La constante de proportionnalité β/T_0^2 est le coefficient thermique de la résistance.

◇

EXEMPLE 1.9 Considérons les ondes se propageant à la vitesse de phase

$$c = \sqrt{\frac{gL}{2\pi} \operatorname{th} \frac{2\pi D}{L}}$$

et considérons cette expression comme une fonction de la profondeur uniquement.

En appliquant les règles usuelles de dérivation[3], on a

$$\frac{dc}{dD} = \sqrt{\frac{g\pi}{2L \ \operatorname{th} \frac{2\pi D}{L}}} \ \frac{1}{\operatorname{ch}^2 \frac{2\pi D}{L}}$$

En particulier, si $D^\star = L/20$, alors

$$\begin{aligned} c(D) &\approx c(D^\star) + c'(D^\star)(D - D^\star) \\ &\approx c(D^\star) + 2.062\sqrt{\frac{g}{L}}(D - D^\star) \end{aligned}$$

En première approximation, toute augmentation de la profondeur de un mètre s'accompagne d'une variation de la vitesse de $2.06\sqrt{g/L}$. ◇

EXEMPLE 1.10 Depuis 1666, l'échelle pratique de salinité est définie par une mesure de conductivité. On considère le ratio K_{15} de la conductivité électrique de l'échantillon d'eau ramené à la température de 15°C et à la pression d'une atmosphère et de la conductivité d'une solution de chlorure de potassium de fraction massique $32.4356\ 10^{-3}$ à la même température et à la même pression. La salinité pratique est alors définie par (UNESCO, 1980)

$$S = 0.0080 - 0.1692K_{15}^{1/2} + 25.3851K_{15} + 14.0941K_{15}^{3/2} - 7.0261K_{15}^2 + 2.7081K_{15}^{5/2}$$

Cette formule est valable dans le domaine de salinité pratique allant de 2 à 42 (Rappelons que la salinité pratique est sans unité).

Par définition, la salinité pratique est égale à 35 pour un rapport K_{15} unitaire.

Le graphe de S est représenté à la figure 1.4. On constate que la salinité pratique varie quasiment linéairement avec K_{15}. Dès lors, si on travaille dans un domaine limité de salinité autour

[3]Rappelons que

$$\frac{d}{dx}\operatorname{sh} x = \operatorname{ch} x, \qquad \frac{d}{dx}\operatorname{ch} x = \operatorname{sh} x \qquad \frac{d}{dx}\operatorname{th} x = \frac{1}{\operatorname{ch}^2 x} = 1 - \operatorname{th}^2 x \qquad (1.32)$$

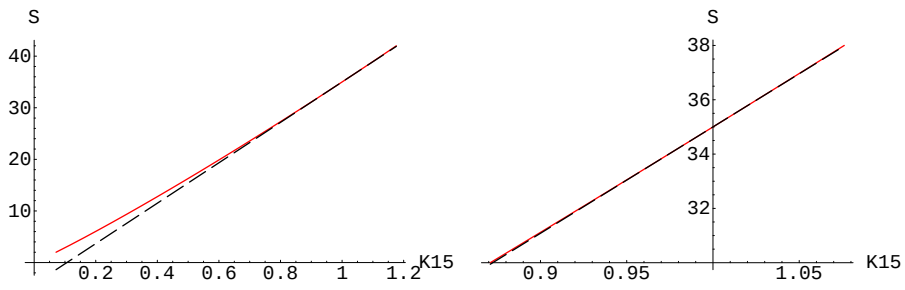


FIG. 1.4 – Salinité pratique en fonction du rapport K_{15}. Loi réelle (trait continu en rouge) et loi linéaire approchée (en pointillé).

de $S = 35$, on peut écrire

$$\begin{aligned} S &\approx S(1) + S'(1)(K_{15} - 1) \\ &\approx 35.00 + 39.1597 * (K_{15} - 1) \\ &\approx 39.1597 K_{15} - 4.1597 \end{aligned}$$

On vérifie sur la figure 1.4 que cette loi linéaire constitue une très bonne approximation de la loi réelle. L'erreur maximale sur la salinité est de 0.05615 si on se limite aux salinité comprises entre 30 et 38 !

◇

Dans le cas d'une fonction de plusieurs variables $f(x_1, x_2, \ldots, x_k, \ldots, x_n)$, on définit la *dérivée partielle par rapport à la variable* x_k comme la dérivée de la fonction d'une variable obtenue en bloquant toutes les variables sauf x_k. On écrira

$$\frac{\partial f}{\partial x_k}(x_1^\star, x_2^\star, \ldots, x_k^\star, \ldots, x_n^\star) = \lim_{x_k \to x_k^\star} \frac{f(x_1^\star, x_2^\star, \ldots, x_k, \ldots, x_n^\star) - f(x_1^\star, x_2^\star, \ldots, x_k^\star, \ldots, x_n^\star)}{x_k - x_k^\star} \tag{1.33}$$

Cette dérivée partielle peut encore être interprétée comme le taux de variation de la grandeur f lorsqu'on fait varier x_k et que l'on bloque toutes les autres variables. Si f est suffisamment régulière[4], on écrira

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) - f(x_1^\star, x_2^\star, \ldots, x_k^\star, \ldots, x_n^\star) = \\ \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x_1^\star, x_2^\star, \ldots, x_k^\star, \ldots, x_n^\star)(x_k - x_k^\star) \\ + o(x_1 - x_1^\star, \ldots, x_n - x_n^\star) \end{aligned} \tag{1.34}$$

[4]En vérité, si f est différentiable au point considéré.

qui généralise (1.31). Avec les notations matricielles

$$\mathsf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \cdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix} \tag{1.35}$$

où ∇f est appelé le *gradient* de f, on a, de façon compacte,

$$f(\mathsf{x}) = f(\mathsf{x}^\star) + \nabla f^{\mathrm{T}}(\mathsf{x}^\star)(\mathsf{x}-\mathsf{x}^\star) + o(\|\mathsf{x}-\mathsf{x}^\star\|) \tag{1.36}$$

où

$$\|\mathsf{x}-\mathsf{x}^\star\| = \sqrt{\sum_{i=1}^{n}(x-x_i^\star)^2} \tag{1.37}$$

désigne la norme de $\mathsf{x}-\mathsf{x}^\star$. De même que la dérivée $f'(x^\star)$ désigne le taux de variation de f lorsque x augmente, le gradient $\nabla f(\mathsf{x}^\star)$ décrit les taux de variations de f lorsque les différentes variables composant x varient indépendamment.

EXEMPLE 1.11 La densité de l'eau de mer est donnée en kg/m^3 par

$$\rho = \frac{\rho_0}{1-p/K}$$

où

$$\begin{aligned}\rho_0 = 999.842594 &+ 6.793952\cdot 10^{-2}t - 9.095290\cdot 10^{-3}t^2 + 1.001685\cdot 10^{-4}t^3 \\ &- 1.120083\cdot 10^{-6}t^4 + 6.536336\cdot 10^{-9}t^5 \\ &+ (8.24493\cdot 10^{-1} - 4.0899\cdot 10^{-3}t + 7.6438\cdot 10^{-5}t^2 - 8.2467\cdot 10^{-7}t^3 + 5.3875\cdot 10^{-9}t^4)S \\ &+ (-5.72466\cdot 10^{-3} + 1.0227\cdot 10^{-4}t - 1.6546\cdot 10^{-6}t^2)S^{3/2} + 4.8314\cdot 10^{-4}S^2\end{aligned} \tag{1.38}$$

et

$$\begin{aligned}K = 19652.21 &+ 148.4206t - 2.327105t^2 + 1.360477\cdot 10^{-2}t^3 - 5.155288\cdot 10^{-5}t^4 \\ &+ S(54.6746 - 0.603459t + 1.09987\cdot 10^{-2}t^2 - 6.1670\cdot 10^{-5}t^3) \\ &- S^{3/2}(7.944\cdot 10^{-2} + 1.6483\cdot 10^{-2}t - 5.3009\cdot 10^{-4}t^2) \\ &+ p[3.239908 + 1.43713\cdot 10^{-3}t + 1.16082\cdot 10^{-4}t^2 - 5.77905\cdot 10^{-7}t^3 \\ &+ S(2.2838\cdot 10^{-3} - 1.0981\cdot 10^{-5}t - 1.6078\cdot 10^{-6}t^2) \\ &+ S^{3/2}(1.91075\cdot 10^{-4})] \\ &+ p^2[8.50935\cdot 10^{-5} - 6.12293\cdot 10^{-6}t + 5.2787\cdot 10^{-8}t^2 \\ &+ S(-9.9348\cdot 10^{-7} + 2.0816\cdot 10^{-8}t + 9.1697\cdot 10^{-10}t^2)]\end{aligned} \tag{1.39}$$

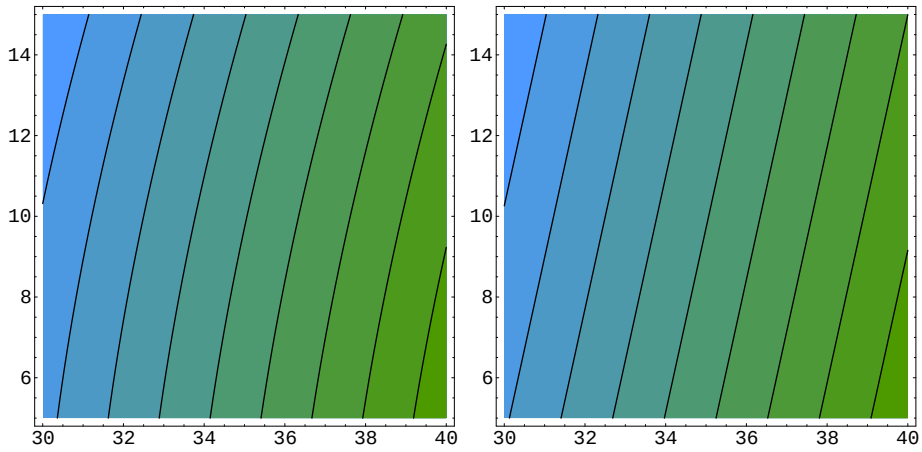


FIG. 1.5 – Masse volumique réelle (à gauche) et approximation linéaire (à droite) en fonction de la température (axe vertical) et de la salinité (axe horizontal). Isovaleurs de 1023 à 1031 (du bleu au vert).

où la température t est exprimée en $^\circ C$ et la pression p en bar.

Sauf si on travaille dans de grandes profondeurs, l'effet de la pression est négligeable. Par exemple, à une température de 10°C et une salinité de 35, la sensibilité à la pression est donnée par

$$\frac{\partial \rho}{\partial p}(t=10, S=35, p=0) = 0.04541 \ \ kg/m^3/bar$$

ce qui signifie qu'il faut augmenter la pression de 2.2 *bars* pour augmenter la densité de 0.1 kg/m^3.

Si on travaille à une température proche de 10°C et une salinité proche de 35, on peut linéariser l'expression complexe de la densité selon

$$\begin{aligned}\rho &\approx \rho(t=10,S=35,p=0) + \frac{\partial \rho}{\partial t}(t=10,S=35,p=0)(t-10) \\ &\qquad + \frac{\partial \rho}{\partial S}(t=10,S=35,p=0)(S-35) \\ &\approx 1026.95 - 0.17129(t-10) + 0.781093(S-35) \\ &\approx 1001.324645 - 0.17129t + 0.781093S\end{aligned}$$

On retrouve, comme attendu, que la densité est une fonction croissante de la salinité et décroissante de la température.

Pour des températures allant de 5 à 15°C et des salinités variant entre 30 et 40, l'erreur maximale associée à cette approximation linéaire est une erreur en excès de 0.1872 kg/m^3 aux extrémités de l'intervalle. ◇

1.3.1 Approximation de Taylor.

Les approximations linéaires fournies par linéarisation peuvent être améliorées et l'erreur peut être quantifiée en utilisant la formule de Taylor.

Dans le cas d'une fonction d'une seule variable, celle-ci s'écrit de la façon suivante. Si la fonction réelle f est n fois continûment dérivable sur un intervalle $[a,x]$ (ou $[x,a]$) et $n+1$ fois dérivable sur l'intervalle ouvert correspondant $]a,x[$ (ou $]x,a[$), alors il existe au moins un point $\xi \in]a,x[$ (ou $\in]x,a[$) tel que

$$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \cdots \\ \cdots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi) \quad (1.40)$$

Le cas particulier où $a = 0$ se rencontre très fréquemment en pratique. La formule de Taylor porte alors le nom de *formule de MacLaurin.*

$$f(x) = f(0) + \frac{x}{1!} f'(0) + \frac{x^2}{2!} f''(0) + \cdots \\ \cdots + \frac{x^n}{n!} f^{(n)}(0) + \frac{x^{n+1}}{(n+1)!} f^{(n+1)}(\theta x) \quad (\theta \in]0,1[) \quad (1.41)$$

Remarquons que le développement de MacLaurin d'une fonction paire ne peut comporter que des puissances paires de x tandis que celui d'une fonction impaire ne présente que des puissances impaires de x. Ceci provient du fait que la dérivée d'une fonction paire est impaire et la dérivée d'une fonction impaire est paire. Or une fonction impaire est nécessairement nulle à l'origine.

L'erreur commise en approchant la fonction réelle par son *développement limité*

$$f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \cdots + \frac{(x-a)^n}{n!} f^{(n)}(a) \quad (1.42)$$

est donné par

$$R_n(x) = (x-a)^{n+1} \frac{f^{(n+1)}(\xi)}{(n+1)!} \quad (1.43)$$

Au moyen d'une majoration appropriée de $f^{(n+1)}$, on peut alors fixer une borne d'erreur et on a

$$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \cdots \\ \cdots + \frac{(x-a)^n}{n!} f^{(n)}(a) + O\big[(x-a)^{n+1}\big], \qquad x \to a \quad (1.44)$$

Si la fonction est suffisamment régulière, la précision du développement limité peut être améliorée en augmentant le nombre de termes. On peut utiliser ce résultat

pour approcher les fonctions transcendantes aussi précisément que nécessaire par des polynômes de degré de plus en plus élevé. Si la fonction f est indéfiniment continûment dérivable, il existe en général un domaine dans lequel la fonction est représentée exactement par une série de puissances, *i.e.* un polynôme de degré infini (Cf. tableau 1.1).

$$\frac{1}{1+x} = 1-x+x^2-x^3+\cdots+(-1)^k x^k+\cdots \qquad -1<x<1$$

$$\frac{1}{1-x} = 1+x+x^2+x^3+\cdots+x^k+\cdots \qquad -1<x<1$$

$$(1+x)^\alpha = 1+\alpha x+\frac{\alpha(\alpha-1)}{2!}x^2+\frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3+\cdots+C_\alpha^k x^k+\cdots \qquad -1<x<1$$

$$\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots+(-1)^k\frac{x^{k+1}}{k+1}+\cdots \qquad -1<x\leq 1$$

$$\ln(1-x) = -x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}-\cdots-\frac{x^{k+1}}{k+1}+\cdots \qquad -1\leq x<1$$

$$\frac{1}{2}\ln\frac{1+x}{1-x} = x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots+\frac{x^{2k+1}}{2k+1}+\cdots \qquad -1<x<1$$

$$\mathrm{e}^x = 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots+\frac{x^k}{k!}+\cdots \qquad \forall\, x$$

$$\mathrm{e}^{-x} = 1-x+\frac{x^2}{2!}-\frac{x^3}{3!}+\cdots+(-1)^k\frac{x^k}{k!}+\cdots \qquad \forall\, x$$

$$\mathrm{ch}\,x = 1+\frac{x^2}{2!}+\frac{x^4}{4!}+\frac{x^6}{6!}+\cdots+\frac{x^{2k}}{(2k)!}+\cdots \qquad \forall\, x$$

$$\mathrm{sh}\,x = x+\frac{x^3}{3!}+\frac{x^5}{5!}+\frac{x^7}{7!}+\cdots+\frac{x^{2k+1}}{(2k+1)!}+\cdots \qquad \forall\, x$$

$$\cos x = 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\cdots+(-1)^k\frac{x^{2k}}{(2k)!}+\cdots \qquad \forall\, x$$

$$\sin x = x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots+(-1)^k\frac{x^{2k+1}}{(2k+1)!}+\cdots \qquad \forall\, x$$

TAB. 1.1

EXEMPLE 1.12 Considérons les fonctions $f(x)=\sin x$ et $g(x)=\cos x$ au voisinage de $x=0$. Ces

fonctions sont indéfiniment continûment dérivables sur $\mathbb{R}$ avec

$$f^{(n)}(x) = \sin\left(x + n\frac{\pi}{2}\right) \qquad \text{et} \qquad g^{(n)}(x) = \cos\left(x + n\frac{\pi}{2}\right)$$

de sorte que

$$f^{(n)}(0) = \begin{cases} 0 & \text{si } n = 2k \\ (-1)^k & \text{si } n = 2k+1 \end{cases}$$

$$g^{(n)}(0) = \begin{cases} (-1)^k & \text{si } n = 2k \\ 0 & \text{si } n = 2k+1 \end{cases}$$

Les dérivées successives de $\sin x$ et $\cos x$ sont bornées par 1 indépendamment de n de sorte que

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

En pratique, on peut tronquer ce développement et n'en utiliser que les quelques premiers termes si on désire utiliser ces développements limités au voisinage de $x = 0$. En effet, si on retient seulement $n+1$ termes

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \frac{x^{2n+3}}{(2n+3)!} f^{(2n+3)}(\theta_1 x)$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \frac{x^{2n+2}}{(2n+2)!} f^{(2n+2)}(\theta_2 x)$$

où θ_1 et $\theta_2 \in]0,1[$, les erreurs sont bornées selon

$$\left| \frac{x^{2n+3}}{(2n+3)!} f^{(2n+3)}(\theta_1 x) \right| \leq \frac{|x|^{2n+3}}{(2n+3)!}$$

et

$$\left| \frac{x^{2n+2}}{(2n+2)!} f^{(2n+2)}(\theta_2 x) \right| \leq \frac{|x|^{2n+2}}{(2n+2)!}$$

Si on désire évaluer $\sin(\pi/6)$ et $\cos(\pi/6)$, par le développement limité de MacLaurin, on obtient

Nombre de termes	1	2	3	4
$\sin(\pi/6)$	0.523599	0.499674	0.500002	0.5
Erreur maximale	0.024	0.00033	$2.1\,10^{-6}$	$8.2\,10^{-9}$
$\cos(\pi/6)$	1.	0.862922	0.866054	0.866025
Erreur maximale	0.14	0.0031	0.000029	$1.4\,10^{-7}$

En pratique, on écrira souvent, avec un degré d'approximation d'autant meilleur que x est petit,

$$\sin x \approx x \qquad \text{et} \qquad \cos x \approx 1 - \frac{x^2}{2}$$

La validité de cette approximation dans une large gamme de valeurs de x est confirmée par l'examen des graphe des fonctions trigonométriques et de leurs approximations (Fig. 1.6).

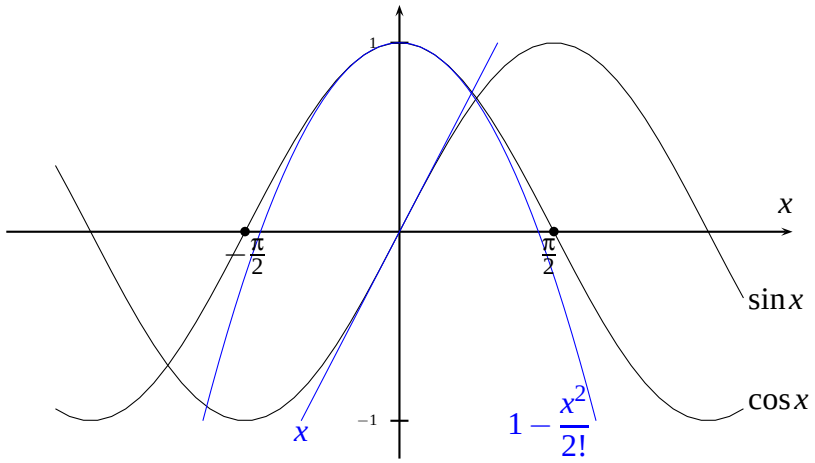


FIG. 1.6

Si x n'est pas petit, on peut obtenir une précision supérieure en ajoutant des termes supplémentaires au développement ou en développant la fonction en série de Taylor autour d'un autre point. ◇

À plusieurs dimensions, la formule de Taylor s'énonce de la façon suivante. Soit Ω un ouvert de $\mathbb{R}^n$ et f une fonction réelle $\in C_p(\Omega)$. Si le segment joignant les points x et $\mathsf{x}+\Delta\mathsf{x}$ est entièrement dans Ω, alors il existe $\theta \in]0,1[$ tel que

$$\begin{aligned} f(\mathsf{x}+\Delta\mathsf{x}) = f(\mathsf{x}) &+ \sum_{i=1}^{n} \Delta x_i \frac{\partial f}{\partial x_i}(\mathsf{x}) \\ &+ \frac{1}{2!} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \Delta x_{i_1} \Delta x_{i_2} \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(\mathsf{x}) + \cdots \\ &+ \frac{1}{p!} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_p=1}^{n} \Delta x_{i_1} \Delta x_{i_2} \ldots \Delta x_{i_p} \frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_p}}(\mathsf{x}+\theta\Delta\mathsf{x}) \end{aligned} \tag{1.45}$$

L'écriture de la formule de Taylor à un ordre quelconque est particulièrement lourde. Dans le cas $p = 2$, le formalisme matriciel permet cependant d'alléger l'écriture. Il vient en effet en introduisant la *matrice hessienne* de f

$$\mathsf{H}(\mathsf{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathsf{x}) \right] \quad (i, j = 1, \ldots, n) \tag{1.46}$$

$$f(\mathrm{x}+\Delta\mathrm{x}) = f(\mathrm{x}) + \Delta\mathrm{x}^{\mathrm{T}}\nabla f(\mathrm{x}) + \frac{1}{2}\Delta\mathrm{x}^{\mathrm{T}}\mathrm{H}(\mathrm{x}+\theta\Delta\mathrm{x})\Delta\mathrm{x} \tag{1.47}$$

où $\Delta\mathrm{x}$ représente la matrice colonne

$$\Delta\mathrm{x} = \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{pmatrix}$$

1.3.2 Différences finies.

Lorsqu'il s'agit de traiter des données expérimentales ou de traiter numériquement des problèmes, les fonctions décrivant les variables dépendantes du problème ne sont pas connues analytiquement et ne peuvent être dérivées en utilisant les règles habituelles de dérivation. La dérivée est alors approchée par un quotient différentiel, appelé *différence finie, i.e.*

$$f'(x^\star) \approx \frac{f(x^\star+\Delta x) - f(x^\star)}{\Delta x} \tag{1.48}$$

qui peut être évalué si les valeurs de f sont connues aux points $x^\star$ et $x^\star + \Delta x$. Cette procédure revient simplement à ignorer la limite dans la définition (1.25) ou encore à remplacer la tangente par la corde dans le graphe de f (Fig. 1.3). Cette façon de procéder est justifiée par la formule de Taylor. En effet, puisque

$$f(x^\star+\Delta x) = f(x^\star) + f'(x^\star)\Delta x + O(\Delta x^2) \tag{1.49}$$

il vient

$$f'(x^\star) = \frac{f(x^\star+\Delta x) - f(x^\star)}{\Delta x} + O(\Delta x) \tag{1.50}$$

L'erreur commise en approchant la dérivée par différence finie tend donc vers zéro linéairement avec l'accroissement Δx. La précision est d'autant plus grande que les données utilisées pour évaluer la dérivée sont proches l'une de l'autre.

Comme le montre (1.50), l'erreur associée à l'approximation (1.48) de la dérivée est du premier ordre en l'accroissement. On peut améliorer ce résultat en remarquant que l'approximation (1.48) revient à évaluer la dérivée en $x^\star$ en explorant le comportement de la fonction uniquement à droite de $x^\star$ (si Δx est positif). On dit que la différence est *décentrée*. L'approximation centrée de la dérivée s'écrit

$$f'(x^\star) \approx \frac{f(x^\star+\Delta x) - f(x^\star-\Delta x)}{2\Delta x} \tag{1.51}$$

On peut vérifier que cette approximation est de meilleure qualité que la précédente en appliquant la formule de Taylor aux deux termes du membre de droite (en supposant f suffisamment régulière) :

$$f(x^\star+\Delta x) = f(x^\star) + f'(x^\star)\Delta x + \frac{1}{2}f''(x^\star)\Delta x^2 + \frac{1}{6}f'''(x^\star)\Delta x^3 + O(\Delta x^4) \tag{1.52}$$

$$f(x^\star-\Delta x) = f(x^\star) - f'(x^\star)\Delta x + \frac{1}{2}f''(x^\star)\Delta x^2 - \frac{1}{6}f'''(x^\star)\Delta x^3 + O(\Delta x^4) \tag{1.53}$$

Substituant ces expressions dans (1.51), il vient

$$\frac{f(x^\star+\Delta x)-f(x^\star-\Delta x)}{2\Delta x} = f'(x^\star)+O(\Delta x^2) \tag{1.54}$$

ce qui montre que l'erreur est maintenant du second ordre en l'accroissement Δx et tend donc vers zéro plus rapidement que dans le cas de l'approximation décentrée.

Si on désire évaluer la dérivée seconde d'une grandeur définie aux noeuds d'un réseau régulier, on écrira encore, en utilisant (1.52)-(1.53),

$$f''(x^\star) = \frac{f(x^\star+\Delta x)-2f(x^\star)+f(x^\star-\Delta x)}{\Delta x^2}+O(\Delta x^2) \tag{1.55}$$

Ces approximations des dérivées sont utilisées pour remplacer les problèmes différentiels continus par des problèmes discrets en vue de leur résolution numérique sur ordinateur.

1.3.3 Dérivée et modélisation.

Comme le montre sa définition, la dérivée est utilisée pour évaluer la sensibilité d'une grandeur, dite *dépendante*, aux variations d'une ou plusieurs autres variables, dites *indépendantes*.

Dans un grand nombre de cas, la variable dépendante n'est pas connue. Seule sa sensibilité aux variables indépendantes (ou l'expression de cette sensibilité en fonction des variables indépendantes) est connue. Dans ce cas, la prise en considération de toutes les sources possibles de variations de la variable dépendante permet d'écrire une équation différentielle pour celle-ci. Les équations de bilan, exprimant le taux de variation de la masse d'une substance en fonction des flux de celle-ci au travers des frontières du système considéré, constituent un cas particulier important.

EXEMPLE 1.13 Exprimons le bilan d'azote dans le bassin occidental de la Mer Méditerranée. Soit M_N la masse totale d'azote. On a

$$\frac{dM_N}{dt} = Flux_{In} - Flux_{Out}$$

où $Flux_{In}$ et $Flux_{Out}$ représentent respectivement le flux total entrant et sortant du basin considéré par unité de temps.

En considérant les principales sources d'échange d'azote entre le système considéré et l'extérieur, on a

$$\frac{dM_N}{dt} = Flux_{Gibraltar} + Flux_{Sicile} + Flux_{Rivieres} + Flux_{Atmosphere} + Flux_{Sediments}$$

où les différents termes du membre de droite expriment respectivement le taux d'échange d'azote au travers des détroit de Gibraltar et de Sicile, les flux apportés par les rivières et l'atmosphère

ainsi que l'échange avec les sédiments. Selon la convention habituelle, tous ces flux sont positifs s'ils contribuent à un apport pour le système et négatifs dans le cas contraire.

Si tous les flux sont connus, l'évolution de la masse de nitrate peut être évaluée en résolvant l'équation différentielle ci-dessus. ◇

EXEMPLE 1.14 Considérons l'intensité lumineuse $I(z)$ régnant dans la colonne d'eau à une profondeur z. La différence entre les intensités à deux profondeurs z et $z+dz$ différentes résulte essentiellement de l'absorbtion du rayonnement lumineux dans la couche d'eau séparant ces deux profondeurs. Si cette absorbtion ne dépend que de l'épaisseur de la couche d'eau et est proportionnelle à celle-ci, on a

$$I(z+dz) - I(z) \approx -k\, I(z)\, dz$$

avec un degré d'approximation d'autant meilleur que l'épaisseur dz est faible. (Remarquez le signe négatif indiquant que l'intensité lumineuse décroît avec la profondeur). À la limite, il vient

$$\lim_{dz\to 0} \frac{I(z+dz)-I(z)}{dz} = \frac{dI}{dz} = -kI(z)$$

On vérifie aisément que cette loi correspond à la loi habituelle

$$I(z) = I(0)\,\mathrm{e}^{-kz}$$

de réduction exponentielle de la lumière avec la profondeur. ◇

1.4 Dérivée d'une fonction composée et dérivée matérielle.

La règle de dérivation des fonctions composées s'énonce comme suit.

Si les fonctions réelles $f_1, f_2, \ldots, f_p$ sont dérivables sur un ouvert Ω de $\mathbb{R}^n$ et si la fonction F est continûment dérivable sur un ouvert ω de $\mathbb{R}^p$ tel que $[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})] \in \omega$ pour tout $\mathsf{x} \in \Omega$, alors pour $k = 1, 2, \ldots, n$, la fonction

$$g(\mathsf{x}) = F[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})]$$

est dérivable par rapport à x_k et on a

$$\frac{\partial g}{\partial x_k}(\mathsf{x}) = \sum_{j=1}^{p} \frac{\partial F}{\partial X_j}[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})] \frac{\partial f_j}{\partial x_k}(\mathsf{x}) \tag{1.56}$$

La difficulté principale pour appliquer la règle de dérivation des fonctions composées réside dans la compréhension de la signification des différents symboles de dérivation et

l'identification des variables indépendantes du problème. En particulier, l'expression

$$\frac{\partial F}{\partial X_j}[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})]$$

représente la dérivée partielle de la fonction F par rapport à sa $j^{\text{ème}}$ variable évaluée au point $[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})] \in \omega$. Par contre, l'expression

$$\frac{\partial g}{\partial x_k}(\mathsf{x}) = \frac{\partial}{\partial x_k} F[f_1(\mathsf{x}), f_2(\mathsf{x}), \ldots, f_p(\mathsf{x})]$$

représente la dérivée partielle de la fonction composée g par rapport à sa $k^{\text{ème}}$ variable et évaluée au point $\mathsf{x} \in \Omega$. Bien que l'écriture des dérivées partielles fasse apparaître le nom d'une variable (X_j ou x_k dans ce qui précède), c'est davantage le numéro de la dérivée partielle et le domaine de définition de la fonction composée qui importent.

EXEMPLE 1.15 Si on néglige l'influence de la pression, la masse par unité de volume de l'eau de mer varie avec la température T et la salinité S selon un loi du type

$$\rho = \rho(T, S)$$

Connaissant les variations de T et de S avec la coordonnée verticale z, on peut calculer le taux de variation correspondant de la densité en considérant

$$\rho = \rho(T(x,y,z,t), S(x,y,z,t))$$

et

$$\frac{\partial \rho}{\partial z} = \frac{\partial \rho}{\partial T}\frac{\partial T}{\partial z} + \frac{\partial \rho}{\partial S}\frac{\partial S}{\partial z}$$

où les dérivées spatiales de T et S sont calculées au point et à l'instant considérés (x,y,z,t) et où les dérivées partielle de ρ par rapport à T et S sont évaluées pour les valeurs de la température et de la salinité effectivement mesurées. ◇

La formule (1.56) doit être adaptée selon les différents cas particuliers rencontrés. Ainsi, les dérivées partielles qui apparaissent dans cette expression seront remplacées par des dérivées habituelles lorsqu'elles s'appliquent à des fonctions d'une seule variable.

La règle de dérivation des fonctions composées permet aussi d'introduire la *dérivée totale* d'une fonction F de plusieurs variables comme la dérivée de la fonction d'une variable réelle t obtenue en explicitant toutes les dépendances des différentes variables de

F par rapport à cette variable. Ainsi, on écrira

$$\begin{aligned}
&\frac{d}{dt}F(f_1(t),\ldots,f_p(t),t)\\
&\qquad=\lim_{\Delta t\to 0}\frac{F(f_1(t+\Delta t),\ldots,f_p(t+\Delta t),t+\Delta t)-F(f_1(t),\ldots,f_p(t),t)}{\Delta t}\\
&\qquad=\sum_{j=1}^{p}\frac{\partial F}{\partial X_j}(f_1(t),\ldots,f_p(t),t)\frac{df_j}{dt}(t)+\frac{\partial F}{\partial t}(f_1(t),\ldots,f_p(t),t) \qquad (1.57)
\end{aligned}$$

EXEMPLE 1.16 Soit un sous-marin se déplaçant dans l'océan. À chaque instant t, les coordonnées du sous-marin sont données par le triplet $(x(t),y(t),z(t))$. Si la distribution de la température est donnée, dans le même système de coordonnées, par la loi $T=T(x,y,z,t)$, alors le thermomètre embarqué à bord du sous-marin indiquera une température T_s telle que

$$T_s(t)=T(x(t),y(t),z(t),t)$$

Le taux de variation temporelle de la température à bord du sous-marin est donc donné par

$$\frac{dT_s}{dt}=\frac{\partial T}{\partial x}\frac{dx}{dt}+\frac{\partial T}{\partial y}\frac{dy}{dt}+\frac{\partial T}{\partial z}\frac{dz}{dt}+\frac{\partial T}{\partial t}$$

où les dérivées partielles de T doivent être évaluées au point $(x(t),y(t),z(t),t)$.

Si on note u,v et w les trois composantes de la vitesse du sous-marin, alors

$$\frac{dx}{dt}=u,\quad \frac{dy}{dt}=v,\quad \frac{dz}{dt}=w$$

et

$$\frac{dT_s}{dt}=\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+w\frac{\partial T}{\partial z}\right)+\frac{\partial T}{\partial t}\neq\frac{\partial T}{\partial t}$$

Les variations de température proviennent donc du mouvement du sous-marin se déplaçant dans un milieu de température inhomogène et des variations temporelles locales de température de l'eau. ◇

1.4.1 Gradient et dérivée directionnelle.

La connaissance des dérivées partielles d'ordre un d'une fonction différentiable suffit à déterminer complètement le taux de variation de la fonction dans n'importe quelle direction. En effet, si on désire évaluer le taux de variation de f dans la direction repérée par le vecteur de composantes $(\cos\alpha,\sin\alpha)$, on forme le quotient différentiel

$$\frac{f(x+\theta\cos\alpha,y+\theta\sin\alpha)-f(x,y)}{\theta}=\cos\alpha\frac{\partial f}{\partial x}(x,y)+\sin\alpha\frac{\partial f}{\partial y}(x,y)+\frac{o(|\theta|)}{\theta}$$

en tenant compte de la différentiabilité de f. Il vient alors

$$\lim_{\theta\to 0}\frac{f(x+\theta\cos\alpha,y+\theta\sin\alpha)-f(x,y)}{\theta}=\frac{\partial f}{\partial x}(x,y)\cos\alpha+\frac{\partial f}{\partial y}(x,y)\sin\alpha \qquad (1.58)$$

qui représente le taux de variation de f dans la direction choisie. Celui-ci est donc obtenu par combinaison linéaire des dérivées partielles de f. En prenant $\alpha=0$ et $\alpha=\pi/2$, on retrouve bien l'interprétation des dérivées partielles de f comme taux de variation de f dans les directions parallèles aux axes.

On peut généraliser ce résultat dans $\mathbb{R}^n$ en adoptant un formalisme vectoriel. Ainsi, on définit le vecteur ∇f (ou $\operatorname{grad} f$), dit *gradient* de f, comme le vecteur de composantes

$$\left(\frac{\partial f}{\partial x_1}(\mathrm{x}),\frac{\partial f}{\partial x_2}(\mathrm{x}),\ldots,\frac{\partial f}{\partial x_n}(\mathrm{x})\right) \qquad (1.59)$$

Le symbole ∇, appelé *nabla*, constitue un *opérateur différentiel vectoriel* qui s'écrit

$$\nabla=\sum_{i=1}^{n}\mathbf{e}_i\frac{\partial}{\partial x_i} \qquad (1.60)$$

dans un système de référence orthonormé. L'application de l'opérateur ∇ à une fonction f donne le gradient de cette fonction.

Si $\mathbf{e}=l_1\mathbf{e}_1+\cdots+l_n\mathbf{e}_n$ désigne un vecteur unitaire, c'est-à-dire tel que

$$\|\mathbf{e}\|=\sqrt{l_1^2+\cdots+l_n^2}=1$$

et si f est différentiable en x, alors

$$\begin{aligned}D_{\mathbf{e}}f(\mathrm{x})&=\lim_{\theta\to 0^+}\frac{f(x_1+\theta l_1,x_2+\theta l_2,\ldots,x_n+\theta l_n)-f(x_1,x_2,\ldots,x_n)}{\theta}\\&=\mathbf{e}\cdot\nabla f\end{aligned} \qquad (1.61)$$

est appelée la *dérivée directionnelle* de f dans la direction du vecteur $\mathbf{e}$.

EXEMPLE 1.17 La dérivée totale introduite dans l'exemple 1.16 peut s'écrire sous la forme

$$\frac{dT}{dt}=\mathbf{v}\cdot\nabla T+\frac{\partial T}{\partial t}$$

◇

À partir de la définition du vecteur gradient, on en déduit que

- Le taux de variation d'une fonction f est maximum dans la direction du vecteur ∇f.
- Le taux de variation de f est minimum dans la direction de $-\nabla f$.
- Le taux de variation de f est nul dans toute direction perpendiculaire à ∇f.

1.5 Primitivation et intégration.

L'introduction de l'intégrale est souvent motivée par le désir de calculer l'aire située sous une courbe $y = f(x)$ dans un intervalle $[a,b]$ (Fig. 1.7).

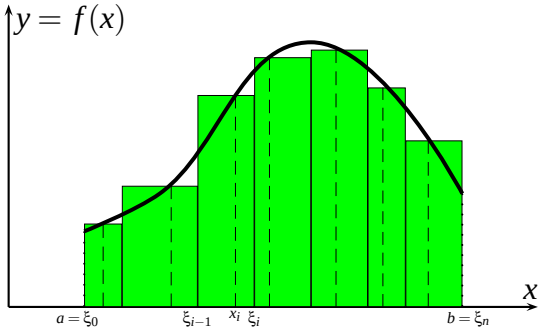


FIG. 1.7

Divisons cet intervalle en n sous-intervalles en introduisant les points intermédiaires ξ_i tels que $a = \xi_0 < \xi_1 < \xi_2 < \ldots < \xi_{n-1} < \xi_n = b$ et formons la somme

$$S_n = \sum_{i=1}^{n} f(x_i)(\xi_i - \xi_{i-1}) \tag{1.62}$$

où x_i est un point arbitraire dans l'intervalle $[\xi_{i-1}, \xi_i]$. Géométriquement, cette somme représente l'aire cumulée de tous les rectangles de la figure 1.7. Si on augmente indéfiniment le nombre de points de la subdivision en faisant tendre n vers l'infini de telle façon que $\Delta\xi = \max_i(\xi_i - \xi_{i-1})$ tende vers zéro, S_n peut avoir ou non une limite unique finie (indépendante du choix de la subdivision et des points d'évaluation de la fonction). Si cette limite existe, on la note

$$\int_a^b f(x)dx = \lim_{n\to+\infty, \Delta\xi\to 0} \sum_{i=1}^{n} f(x_i)(\xi_i - \xi_{i-1}) \tag{1.63}$$

qui s'appelle l'*intégrale définie* (ou plus simplement l'*intégrale*) de $f(x)$ entre a et b. On dit alors que la fonction f est *intégrable au sens de Riemann* dans [a,b]. On dit aussi que $f(x)$ est l'*intégrand*, que $[a,b]$ est le *domaine d'intégration* et que a et b sont les *limites ou bornes d'intégration.*

Notons que l'interprétation géométrique de l'intégrale définie par (1.63) comme surface située entre le graphe de f et l'axe des x n'est exacte que si la fonction est partout positive. Si $f(x)$ prend des valeurs positives et négatives, l'intégrale représente la somme algébrique des aires au-dessus et en-dessous de l'axe des x en considérant comme positives les aires au-dessus de l'axe et négatives celles en-dessous de l'axe des x.

Remarquons encore que, lorsque l'on désire calculer numériquement la valeur approchée d'une intégrale, on renverse généralement la définition (1.63) et on approche la valeur de l'intégrale par une somme finie de termes semblables à ceux apparaissant dans cette expression.

Il va de soi que la définition (1.63) peut être appliquée sans rapport apparent avec la recherche de l'aire définie par le graphe de f. Toute quantité qui peut être exprimée sous la forme de la limite d'une somme comme (1.63) peut être représentée par une intégrale. Ceci correspond à l'approche, très répandue dans les différents domaines de mathématiques appliquées, consistant à décomposer un processus ou un milieu matériel en un très grand nombre d'éléments de tailles très petites et à définir la résultante comme étant la somme sur tous ces petits éléments. Si la taille des éléments tend vers zéro, on dit qu'ils sont *infinitésimaux* et la somme devient une intégrale. Cependant, la quantité ainsi définie (si elle est réelle) pourra aussi être interprétée géométriquement comme l'aire sous un graphe.

EXEMPLE 1.18 On note $TProd(t)$, le taux de production primaire en fonction du temps t. La production primaire totale au cours d'une période allant de t_1 à t_2 est donnée par l'intégrale

$$\int_{t_1}^{t_2} TProd(t)dt$$

◇

EXEMPLE 1.19 Selon la théorie de la *profondeur critique* introduite par Sverdrup, un bloom phytoplanctonique se produit lorsque la profondeur de la couche de mélange est inférieure à la profondeur critique au-dessus de laquelle la production nette est positive.

Le taux de croissance du phytoplancton, *i.e.* le taux d'augmentation de la masse du compartiment phytoplanctonique, est donné par

$$\frac{1}{P}\frac{dP}{dt} = \text{Taux brut de photosynthèse} - \text{Taux de respiration}$$

Dans cette expression, le taux de respiration (dans lequel Sverdrup incorpore également le broutage par les niveaux trophiques supérieurs) peut être considéré comme à peu près indépendant de la profondeur. Le taux de photosynthèse dépend par contre de l'intensité lumineuse et est donc une fonction décroissante de la profondeur. Si on suppose le taux de photosynthèse proportionnel à l'intensité lumineuse, celui-ci décroît exponentiellement.

La profondeur de compensation est le niveau vertical pour lequel les taux bruts de photosynthèse et de respiration sont égaux. Au-dessus de ce niveau, on assiste à une croissance des niveaux phytoplanctoniques. En-dessous, par contre, la respiration dépasse la photosynthèse.

Si le mélange vertical est actif, cependant, la production nette dans la couche de mélange est donnée par l'intégrale sur cette couche. La profondeur critique Z_c est donc définie par la relation

$$\int_0^{Z_c} \frac{1}{P}\frac{dP}{dt}dz = \int_0^{Z_c} \Big(\text{Taux brut de photosynthèse} - \text{Taux de respiration}\Big)dz = 0$$

Sur la couche ainsi définie, la photosynthèse totale est exactement compensée par la respiration.

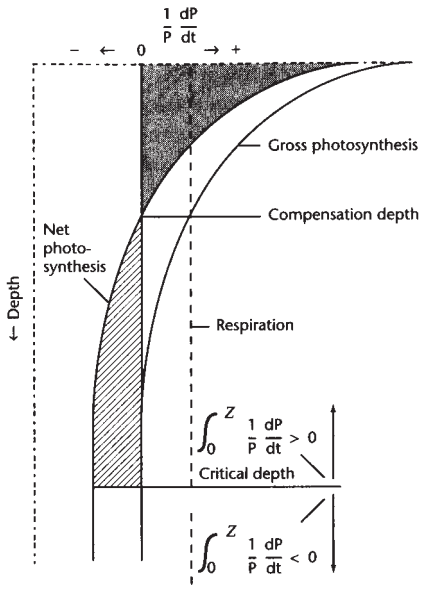


FIG. 1.8

Si on suppose le taux brut de production par photosynthèse strictement proportionnel à l'intensité lumineuse et que celle-ci décroît exponentiellement selon la loi de Beer

$$I(z) = I_0\,\mathrm{e}^{-kz}$$

où k désigne le coefficient d'extinction lumineuse, la profondeur critique Z_c peut être explicitée de la façon suivante. On a

$$\int_0^{Z_c} \left(\alpha I_0\,\mathrm{e}^{-kz} - r\right) dz = 0$$

où r désigne le taux de respiration (constant) et α est le taux de photosynthèse spécifique. En évaluant l'intégrale, la profondeur critique Z_c apparaît comme la solution de l'équation transcendante

$$\frac{1}{k}\alpha I_0(1 - \mathrm{e}^{-kZ_c}) = rZ_c$$

Une solution approchée peut être obtenue en supposant $kZ_c \ll 1$ justifiant l'approximation

$$\mathrm{e}^{-kZ_c} \approx 1 - kZ_c + \frac{1}{2}k^2 Z_c^2$$

Il vient alors

$$\alpha I_0(2 - kZ_c) = 2r$$

soit

$$Z_c \approx \frac{2}{k}\left(1 - \frac{r}{\alpha I_0}\right)$$

(Remarquons que l'hypothèse $kZ_c \ll 1$ est vérifiée si αI_0 est proche de r.) Comme attendu, la profondeur critique décroît avec r et augmente avec I_0.

◇

1.5.1 Moyenne et moyenne glissante.

La moyenne d'une grandeur est également définie par une intégrale. Ainsi, la moyenne de $f(t)$ sur l'intervalle $[0,T]$, notée $< f >$ ou $\bar{f}$ est donnée par

$$\frac{1}{T}\int_0^T f(t)dt \tag{1.64}$$

Cette expression est une généralisation évidente de la définition classique de la moyenne

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.65}$$

d'un ensemble $\{x_1, x_2, \ldots, x_n\}$ de données discrètes. En effet, en présence d'une fonction f continue, la définition (1.63) de l'intégrale revient à remplacer la distribution continue par une distribution discrète matérialisée par chacun des petits rectangles de la figure 1.7. Dans le cas d'une partition de l'intervalle $[0,T]$ en sous-intervalles de même largeur $\Delta x = \xi_i - \xi_{i-1}$, on a

$$\frac{1}{T}\int_0^T f(t)dt = \lim_{n\to+\infty,\Delta x\to 0} \frac{1}{n\Delta x}\sum_{i=1}^{n} f(x_i)\Delta x = \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

en tenant compte de $n\Delta x = T$.

EXEMPLE 1.20 La distribution des caractéristiques d'une population est souvent décrite par une fonction de distribution. Ainsi, la distribution des âges peut être décrite par une fonction $f(a)$ telle que

$$\int_{a_1}^{a_2} f(t)dt$$

donne le nombre d'individus dont l'âge est compris dans l'intervalle $[a_1, a_2]$. Si a_1 et a_2 sont très proches l'un de l'autre, on peut également exprimer cette propriété en disant que $f(a)da$ représente le nombre d'individus dont les âges sont compris entre a et $a+da$ (où da est supposé très petit).

La population totale est donnée par

$$N = \int_0^{a_{max}} f(a)da$$

où a_{max} désigne l'âge maximum (ou toute valeur au-delà de laquelle $f(a)$ s'annule identiquement). L'âge moyen de la population est donné par

$$\bar{a} = \frac{1}{N}\int_0^{a_{max}} f(a)ada = \frac{\int_0^{a_{max}} f(a)ada}{\int_0^{a_{max}} f(a)da}$$

◇

Le calcul de la moyenne est une opération linéaire, *i.e.* quelles que soient les constantes α, β et les fonctions f et g, on a

$$< \alpha f(t) + \beta g(t) >= \alpha < f(t) > + \beta < g(t) > \tag{1.66}$$

Par contre,

$$< f(g(x)) > \neq f(< g(x) >) \tag{1.67}$$

sauf si f est elle-même une fonction linéaire. Ainsi, l'effet moyen d'un forçage agissant sur un système non linéaire n'est pas égal à l'effet du forçage moyen sur ce même système.

EXEMPLE 1.21 Le taux de photosynthèse (par unité de biomasse phytoplanctonique) μ varie avec l'intensité lumineuse I. Si on ignore le phénomène de photoinhibition, la relation $\mu - I$ est caractérisée par une croissance quasi-linéaire de μ pour les faibles intensités lumineuses et l'existence d'un palier μ_{max} aux fortes intensités. Cette relation peut donc être décrite par une loi semblable à celle de Michaelis-Menten (en général, on lui préfère cependant une loi en tangente hyperbolique ou une combinaison d'exponentielles ; la forme retenue ici présente l'avantage de permettre un raisonnement analytique pour illustrer notre propos.)

$$\mu(I) = \mu_{max} \frac{I}{\alpha + I}$$

Considérons une cellule phytoplanctonique qui, en raison du mélange vertical, passe un temps égal à toutes les profondeurs de la couche de mélange d'épaisseur H. Le taux de croissance moyen est donné par

$$< \mu >= \frac{1}{H} \int_0^H \mu_{max} \frac{I(z)}{\alpha + I(z)} dz$$

où $I(z) = I_0 \exp(-kz)$ désigne l'intensité lumineuse à la profondeur z. On calcule aisément

$$\begin{aligned} < \mu > &= \frac{1}{H} \int_0^H \mu_{max} \frac{I_0 \, \mathrm{e}^{-kz}}{\alpha + I_0 \, \mathrm{e}^{-kz}} dz \\ &= \frac{\mu_{max}}{kH} \Big[-\ln(\alpha + I_0 \, \mathrm{e}^{-kz}) \Big]_0^H \\ &= \frac{\mu_{max}}{kH} \ln \left[\frac{\alpha + I_0}{\alpha + I_0 \, \mathrm{e}^{-kH}} \right] \end{aligned}$$

D'autre part, l'intensité lumineuse moyenne sur la couche d'eau considérée est donnée par

$$< I(z) >= \frac{1}{H} \int_0^H I(z) dz = \frac{I_0}{kH} (1 - \mathrm{e}^{-kH})$$

et

$$\mu(< I(z) >) = \frac{\mu_{max} I_0 (1 - \mathrm{e}^{-kH})}{kH \left[\alpha + I_0 (1 - \mathrm{e}^{-kH}) \right]}$$

On constate donc que

$$< \mu(I(z)) > \neq \mu(< I(z) >)$$

i.e. la croissance sous une lumière moyenne n'est pas égale à la croissance moyenne sous un éclairement variable.

Remarquons que c'est la non-linéarité de la relation $\mu - I$ qui est à l'origine de cette différence. Si cette relation était linéaire, *i.e.* du type

$$\mu(I) = \mu_{max}\frac{I}{\alpha}$$

on aurait simplement

$$< \mu(I(z)) >= \mu(< I(z) >) = \frac{\mu_{max} I_0 (1 - \mathrm{e}^{-kH})}{kH\alpha}$$

ce qui peut s'obtenir à partir des résultats précédents en considérant le comportement asymptotique des différentes expressions pour $I_0/\alpha \to 0$ (ce qui revient à considérer que la lumière est toujours insuffisante pour induire un effet de saturation). ◇

La définition (1.64) de la moyenne peut également être utilisée pour calculer une *moyenne glissante* permettant le filtrage rapide des oscillations présentes dans une série temporelle. Il suffit pour ce faire de remplacer la série de départ $f(t)$ par

$$< f(t) >= \frac{1}{T}\int_{t-T/2}^{t+T/2} f(u)du \tag{1.68}$$

L'effet de ce *filtre* est de diminuer fortement les oscillations de période bien inférieure à T et de laisser pratiquement inchangés les signaux de période supérieure à T. En effet, considérons simplement le signal périodique (selon la théorie de Fourier, un signal périodique peut être décomposé en une série de signaux harmoniques dont les pulsations sont des multiples de la pulsation du signal initial)

$$f(t) = A\sin\frac{2\pi t}{\tau} \tag{1.69}$$

Il vient

$$< f(t) >= \frac{\tau A}{2\pi T}\Big[-\cos\frac{2\pi t}{\tau}\Big]_{t-T/2}^{t+T/2} = \frac{A\tau}{T\pi}\sin\frac{\pi T}{\tau}\sin\frac{2\pi t}{\tau} = \frac{\tau}{\pi T}\sin\frac{\pi T}{\tau}f(t) \tag{1.70}$$

En appliquant un tel filtre, on peut ainsi débarrasser (grossièrement) un signal du bruit à haute-fréquence pour se concentrer sur les signaux de plus basse fréquence. Pour que cette opération ait un sens, cependant, il convient de choisir un temps T permettant un partage clair des signaux de hautes et basses fréquences. Le temps T doit donc être strictement compris entre deux temps caractéristiques du système étudié.

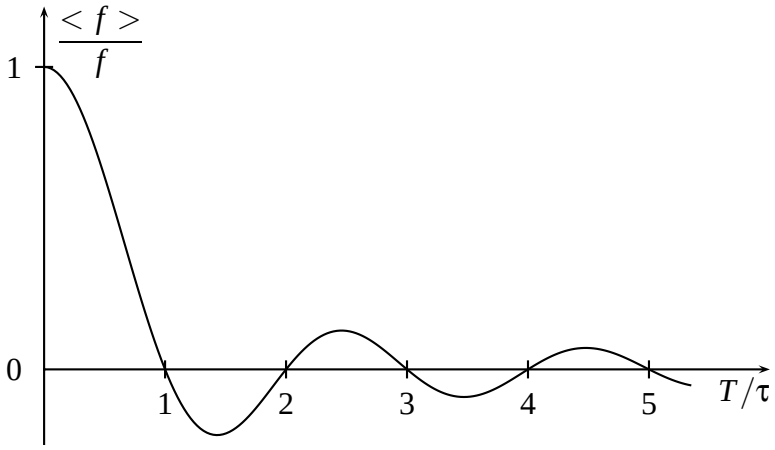


FIG. 1.9 – Influence du filtrage en fonction du rapport des périodes du filtre et du signal initial.

EXEMPLE 1.22 Considérons un signal saisonnier (les variations de la température par exemple) auquel se superpose des oscillations de hautes fréquences correspondant aux variations journalières (T=1 jour) et aux perturbations induites par l'alternance des dépressions des anticyclones Atlantiques (T≈ 8 jours). La figure 1.22 montre le signal brut tel qu'il est enregistré par les capteurs.

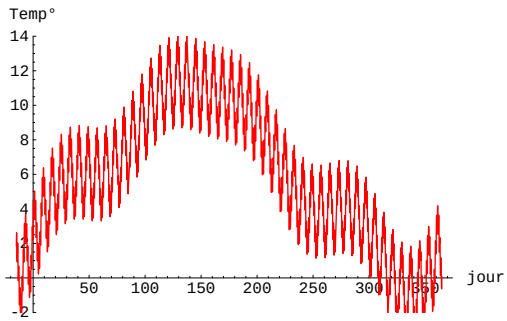


FIG. 1.10 – Signal brut.

Les moyennes glissantes avec un temps caractéristique de 3 jours et de 10 jours ne permettent pas d'isoler le signal saisonnier et ne permettent pas non plus d'éliminer correctement l'influence des dépressions et anticyclones.

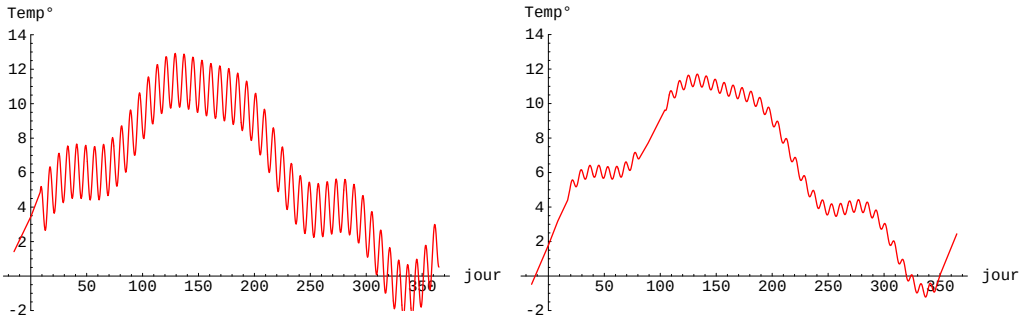


FIG. 1.11 – Moyenne glissante avec T=3 jours (à gauche) et T=10 jours (à droite).

Une moyenne glissante calculée sur une période de 40 jours permet par contre d'éliminer les oscillations non désirées sans affecter exagérément le signal saisonnier.

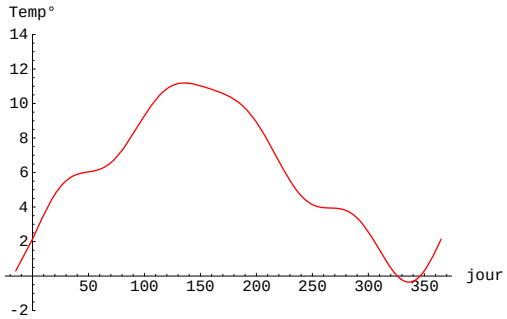


FIG. 1.12 – Moyenne glissante avec T=40 jours.

◇

1.5.2 Primitive.

On appelle primitive d'une fonction continue f sur $\mathtt{I}$, toute fonction F continûment dérivable telle que

$$F'(x) = f(x) \qquad \forall x \in \mathtt{I} \tag{1.71}$$

On montre que toutes les primitives de f ne diffèrent que par une constante additive et sont données par

$$F(x) = \int_{x_0}^{x} f(t)dt + F(x_0) \tag{1.72}$$

où $x_0 \in \mathtt{I}$. Ceci relie la notion de primitive à celle d'intégrale ; la connaissance d'une primitive quelconque F de f permet le calcul des intégrales de f par variation de F, *i.e.*

$$\int_a^b f(x)dx = \Big[F[x]\Big]_a^b = F(b) - F(a) \tag{1.73}$$

D'après la définition (1.71), la primitivation apparaît comme l'opération inverse de la dérivation. Pour toute fonction continue f, on a en effet,

$$\frac{d}{dx}\int_{x_0}^{x} f(t)dt = f(x) \tag{1.74}$$

Plus généralement, on montre (sous certaines conditions de régularité de $a(x)$, $b(x)$ et $f(t,x)$,

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t,x)dt = f(a(x),x)\,a'(x) - f(b(x),x)\,b'(x) + \int_{a(x)}^{b(x)} \frac{\partial f}{\partial x}(t,x)dt \tag{1.75}$$

Chapitre 2

Analyse dimensionnelle.

2.1 Dimensions.

Lorsque l'on construit un modèle d'un système quelconque, on caractérise celui-ci au moyen d'un certain nombre de grandeurs qui décrivent différents aspects de ce système : température, salinité, concentration en éléments nutritifs, éclairement...

Dès lors que l'on désire combiner ces grandeurs dans une même équation, il convient d'être attentif aux unités dans lesquelles ces grandeurs sont exprimées. Plus fondamentalement encore, il est nécessaire de percevoir la nature des grandeurs utilisées : longueur, temps, masse,... Cette nature transparaît au travers des unités utilisées. Différentes unités peuvent cependant être utilisées pour mesurer un même type de grandeurs. Ainsi, le mètre, le pouce ou l'Angstrom sont des unités de mesure des longueurs.

Il apparaît que toutes les grandeurs utilisées en physique, en chimie, en écologie,... font intervenir sept *grandeurs fondamentales* : la masse, la longueur, le temps, la température, le courant électrique, la quantité de matière et l'intensité lumineuse. Le tableau 2.1 présente les unités de base[1] correspondantes dans le Système International d'Unités (SI). Les trois grandeurs fondamentales M, L et T sont suffisantes pour décrire la mécanique Newtonienne. La température [2] θ, la quantité de matière N et l'intensité lumineuse doivent être prises en compte en écologie.

Toutes les grandeurs qui n'apparaissent pas dans le tableau 2.1 sont appelées des *grandeurs dérivées.* Les *dimensions* d'une variable X quelconque sont les produits des puissances des dimensions des grandeurs fondamentales composant cette variable. Ainsi, puisqu'une surface s'exprime comme le produit de deux longueurs, les dimensions d'une surface sont L^2. De même, une vitesse a les dimensions LT^{-1} puisqu'elle exprime l'espace parcouru par unité de temps.

Plus précisément, les dimensions $[X]$ d'une variable X sont caractérisées entièrement

[1] En plus des unités de base, on introduit également les unités supplémentaires que sont le radian (rad) et le steradian (sr) qui mesurent les angles plans et solides.

[2] Remarquons que la température est généralement mesurée en degrés Celsius (°C) dans la plupart des études environnementales.

Grandeur fondamentales	Dimension	Unité de base	Symbole des unités
masse	M	kilogramme	kg
longueur	L	mètre	m
temps	T	seconde	s
courant électrique	I	ampère	A
température	θ	kelvin	K
quantité de matière	N	mole	mol
intensité lumineuse	J	candela	cd

TAB. 2.1 – Grandeurs fondamentales et unités de base associées dans le Système International.

par la donnée des exposants caractéristiques $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta$ tels que

$$[X] = \mathrm{M}^{\alpha}\,\mathrm{L}^{\beta}\,\mathrm{T}^{\gamma}\,\mathrm{I}^{\delta}\,\theta^{\varepsilon}\,\mathrm{N}^{\zeta}\,\mathrm{J}^{\eta} \tag{2.1}$$

Ces exposants caractéristiques déterminent la façon dont la mesure d'une grandeur est affectée par un changement d'unités. Ainsi, si on passe du Système International au système CGS utilisant le centimètre comme unité de base pour la mesure des longueurs, la valeur numérique des longueurs est multipliée par 100 tandis que celle des surfaces (dimensions L^2) est multipliée par 100^2 soit 10 000.

Lorsque tous les exposants caractéristiques d'une grandeur sont égaux à zéro, cette grandeur est dite *adimensionnelle*. Il en est ainsi des angles[3], de la densité relative et de tout autre rapport de grandeurs de dimensions identiques. Les grandeurs adimensionnelles ne sont pas affectées par un changement de système d'unités. Tous les arguments des fonctions transcendantes $(\sin, \cos, \exp, \log, \ldots)$ ainsi que les exposants sont toujours adimensionnels. Les nombres purs $(2, 7, \pi, \mathrm{e}, \ldots)$ sont également adimensionnels.

2.2 Homogénéité dimensionnelle et équation aux dimensions.

La multiplication (resp. la division) de grandeurs de dimensions différentes permet de définir de nouvelles grandeurs, avec des dimensions nouvelles qui sont le produit (resp. le quotient) des dimensions des grandeurs initiales.

EXEMPLE 2.1 Les dimensions du travail mécanique, produit de la force et du déplacement, sont obtenues en multipliant les dimensions d'une force, MLT^{-2}, par celle d'un déplacement, L. Le

[3]Un angle plan est le rapport de la longueur de l'arc intercepté par l'angle et du rayon du cercle portant cet arc. Un angle est donc adimensionnel mais pas sans unités. Selon le SI, sa mesure s'exprime en radians.

travail, comme l'énergie possède donc les dimensions $\mathrm{ML}^2\mathrm{T}^{-2}$. Le flux de chaleur, c'est-à-dire le flux d'énergie thermique s'écoulant par unité de surface et par unité de temps a les dimensions

$$\frac{ML^2T^{-2}}{L^2 \cdot T} = MT^{-3} \tag{2.2}$$

◇

Du point de vue des dimensions, l'intégration et la dérivation par rapport à une variable s'assimilent à la multiplication et à la division.

EXEMPLE 2.2 Si $v(t)$ désigne la vitesse d'un point matériel en fonction du temps, on a

$$\left[\int_{t_0}^{t} v(\tau)d\tau\right] = [v][t] = \mathrm{LT}^{-1}\mathrm{T} = \mathrm{L} \tag{2.3}$$

et

$$\left[\frac{d}{dt}v(t)\right] = \frac{[v]}{[t]} = \frac{\mathrm{LT}^{-1}}{\mathrm{T}} = \mathrm{LT}^{-2} \tag{2.4}$$

◇

L'addition, la soustraction et l'égalité ne sont par contre possibles qu'entre des grandeurs possédant les mêmes dimensions. Si

$$a+b+c+\cdots = g+h+\cdots \tag{2.5}$$

alors, les variables a, b, c,..., g, h, ... doivent toutes avoir les mêmes dimensions. C'est le principe de l'*homogénéité dimensionnelle*. Interprétant les dimensions comme la sensibilité au changement de système d'unités, on peut assimiler (et justifier) ce principe à l'expression de l'indépendance des lois naturelles par rapport au système d'unités utilisé pour décrire le système.

Pour construire des équations bien formées, il faut veiller à respecter le principe d'homogénéité dimensionnelle. Inversement, les dimensions d'une grandeur quelconque X peuvent souvent être obtenues en exprimant l'égalité des dimensions des deux membres d'une équation dans laquelle cette grandeur intervient et en résolvant l'équation obtenue par rapport à $[X]$. Une telle équations est appelée une *équation aux dimensions*.

EXEMPLE 2.3 Si on décrit la croissance d'un animal par une loi

$$\frac{dW}{dt} = R - T \tag{2.6}$$

où W désigne la masse de l'animal, R l'apport alimentaire et T la consommation par son métabolisme, les deux termes du membre de droite devront avoir les mêmes dimensions que le membre de gauche, soit MT^{-1}. ◇

EXEMPLE 2.4 Déterminons les dimensions du coefficient de diffusion de la chaleur κ défini comme l'opposé du coefficient de proportionnalité entre le gradient de la température et le flux de chaleur J,

$$\mathbf{J} = -\kappa \nabla T \tag{2.7}$$

En considérant les dimensions des deux membres de cette équation, on a

$$[\mathbf{J}] = [\kappa][\nabla T] \qquad \text{soit} \qquad MT^{-3} = [\kappa]\theta L^{-1} \tag{2.8}$$

et donc

$$[\kappa] = ML\theta^{-1}T^{-3} \tag{2.9}$$

$\diamond$

2.3 Théorème Pi.

L'analyse des dimensions des paramètres intervenant dans un problème permet de dégager des conclusions rapides concernant l'influence des différents paramètres. En effet, le comportement d'un système ne peut dépendre de la valeur d'un paramètre dimensionnel ; un changement d'unités induirait alors une modification du comportement de ce système. Pour garantir l'indépendance par rapport au système d'unités, les caractéristiques d'un système ne peuvent dépendre que de combinaisons adimensionnelles des paramètres. C'est l'essence du théorème Pi ou théorème de Vashi-Buckingham :

> *Toute équation homogène du point de vue des dimensions peut être transformée en une relation entre les membres d'une famille complète de produits adimensionnels. Si le nombre de paramètres dimensionnels de l'équation initiale est n et si ces paramètres font intervenir N dimensions fondamentales, alors le nombre de produits adimensionnels est égal à* $n-N$.

Le théorème Pi est à la base des essais effectués sur des modèles réduits par les ingénieurs : tant que le prototype et le modèle réduit partagent les mêmes nombres sans dimensions caractéristiques du problème, les résultats mesurés sur le modèle réduit peuvent être extrapolés au prototype.

Le théorème peut être utilisé également pour deviner la forme des lois gouvernant un système ou pour en simplifier la présentation et l'analyse.

EXEMPLE 2.5 Considérons la force exercée sur un courantomètre plongé dans le courant. Une analyse rapide du problème laisse deviner une dépendance de la force F en la dimension D du courantomètre, la densité de l'eau ρ, la vitesse du courant V et la viscosité dynamique de l'eau η, soit

$$F = f(V, D, \rho, \eta) \tag{2.10}$$

Les dimensions des différentes variables du problème sont

$$[F] = \mathrm{MLT}^{-2}, \quad [V] = \mathrm{LT}^{-1}, \quad [D] = \mathrm{L}, \quad [\rho] = \mathrm{ML}^{-3}, \quad [\eta] = \mathrm{ML}^{-1}\mathrm{T}^{-1} \tag{2.11}$$

Elles dépendent des trois dimensions fondamentales M, L et T. Par le théorème Pi, la relation (2.10) entre les 5 variables dimensionnelles peut prendre la forme d'une équation entre 5-3 = 2 produits adimensionnels. Le passage de 5 à 2 variables simplifie évidemment l'analyse du problème. Si on prend le *nombre de Reynolds*

$$Re = \frac{VD\rho}{\eta}, \qquad [Re] = \frac{\mathrm{LT}^{-1} \cdot \mathrm{L} \cdot \mathrm{ML}^{-3}}{\mathrm{ML}^{-1}\mathrm{T}^{-1}} = 1 \tag{2.12}$$

et le *nombre de Newton*

$$Ne = \frac{F}{\rho D^2 V^2}, \qquad [Ne] = \frac{\mathrm{MLT}^{-2}}{\mathrm{ML}^{-3} \cdot \mathrm{L}^2 \cdot \mathrm{L}^2\mathrm{T}^{-2}} = 1 \tag{2.13}$$

la relation (2.10) peut s'exprimer sous la forme

$$Ne = Cd(Re) \tag{2.14}$$

soit

$$F = Cd(Re)\rho V^2 D^2 \tag{2.15}$$

où $Cd(Re)$ est une fonction du nombre de Reynolds qui pourra être déterminée au moyen de mesures en laboratoire. Ces mesures peuvent être réalisées dans les conditions les plus favorables d'un point de vue expérimental, *i.e.* en choisissant librement la vitesse V et la densité ρ du fluide utilisé dans l'expérience. Les résultats obtenus sont valables dans toutes les conditions, quelles que soient la densité ρ, la viscosité dynamique η ou même la dimension D du courantomètre. ◇

Une connaissance a priori du problème est nécessaire pour bien utiliser et exploiter la puissance de l'analyse dimensionnelle. L'identification correcte des paramètres et constantes dimensionnelles à inclure dans l'analyse est capitale. Si des variables importantes sont absentes, les résultats peuvent se révéler incomplets ou même erronés. D'autre part, la solution peut être parasitée et inutilement compliquée par la prise en compte de trop de paramètres.

La présentation de données sous forme adimensionnelle permet également de produire des graphiques beaucoup plus compacts et plus simples à lire.

EXEMPLE 2.6 Considérons le graphique présentant la distribution spatiale de la concentration d'un traceur passif dans un écoulement unidimensionnel caractérisé par une vitesse constante u du courant et un coefficient de diffusion κ lui aussi constant. Pour un rejet initial donné du traceur passif, la concentration $C(t,x)$ dépend du temps t et de la coordonnée spatiale x. Si on considère différentes situations correspondant à différentes vitesses u et/ou coefficients de diffusion κ, l'advection et la diffusion du traceur seront modifiées et la concentration sera affectée. Pour décrire ces différentes situations, il est a priori nécessaire de tracer des courbes $C(t,\cdot)$ pour différentes valeurs de t, de u et de κ. Pour simplifier la présentation graphique et l'analyse des résultats, on peut cependant avoir recours à une approche adimensionnelle en introduisant les

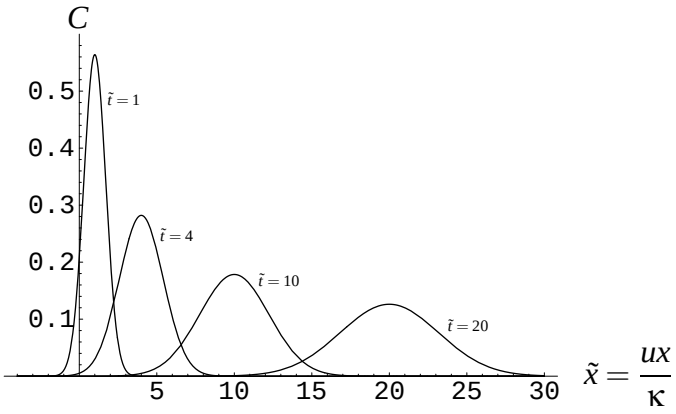


FIG. 2.1 – Distribution de la concentration issue d'un rejet ponctuel en fonction de variables adimensionnelles $\tilde{x}$ pour des temps adimensionnels successifs $\tilde{t} = 1$, 4, 10 et 20.

variables adimensionnels $\tilde{t}$ et $\tilde{x}$ obtenues en divisant t et x par des grandeurs caractéristiques de mêmes dimensions, soit

$$\tilde{t} = \frac{u^2 t}{\kappa}, \qquad \tilde{x} = \frac{xu}{\kappa} \tag{2.16}$$

Puisque x, t, u et κ dépendent uniquement des deux dimensions fondamentales L et T, le champ de concentration peut être décrit comme une fonction des 4-2=2 variables adimensionnelles (2.16). Les courbes correspondantes représentées à la figure 2.1 permettent d'obtenir la solution en tout point x et en tout temps t pour des valeurs quelconques de u et κ.

◇

2.4 Variations caractéristiques.

Les grandeurs adimensionnelles peuvent aussi être introduites par le biais de grandeurs caractéristiques.

EXEMPLE 2.7 Considérons la dynamique d'une population dont la concentration $C(t,x)$ dans un domaine unidimensionnel varie selon

$$\frac{\partial C}{\partial t} = \kappa \frac{\partial^2 C}{\partial x^2} + \mu\, C \tag{2.17}$$

où x désigne la coordonnée spatiale, t est le temps, le premier terme du membre de droite représente le processus de diffusion (κ est le coefficient de diffusion) et le second terme modélise la croissance de la population (μ est le taux de croissance spécifique, supposé constant). La diffusion agit à l'encontre de la croissance de la population. La croissance exponentielle de la population est contrecarrée par la diffusion de celle-ci dans tout l'espace.

Remplaçons les variables dimensionnelles C, t, et x par les variables adimensionnelles

$$\tilde{C} = \frac{C}{C_\star}, \qquad \tilde{t} = \frac{t}{t_\star}, \qquad \tilde{x} = \frac{x}{x_\star} \tag{2.18}$$

où $C_\star$, $t_\star$ et $x_\star$ désignent des valeurs caractéristiques de C, t, et x. Substituant ces expressions dans (2.17), on a

$$\frac{C_\star}{t_\star}\frac{\partial \tilde{C}}{\partial \tilde{t}} = \kappa \frac{C_\star}{x_\star^2}\frac{\partial^2 \tilde{C}}{\partial \tilde{x}^2} + \mu\, C_\star\, \tilde{C} \tag{2.19}$$

Les différents termes de cette expression peuvent être rendus adimensionnels par multiplication par $x_\star^2/(C_\star \kappa)$, soit

$$\left[\frac{x_\star^2}{K t_\star}\right]\frac{\partial \tilde{C}}{\partial \tilde{t}} = \frac{\partial^2 \tilde{C}}{\partial \tilde{x}^2} + \left[\frac{\mu x_\star^2}{\kappa}\right]\tilde{C} \tag{2.20}$$

où les termes entre crochets sont des produits adimensionnels. La dynamique de la population $\tilde{C}$ dépend uniquement de ces deux produits adimensionnels.

En utilisant le second produit adimensionnel, on constate que la longueur caractéristique $x_\star$ est donnée par

$$x_\star \propto \sqrt{\frac{\kappa}{\mu}} \tag{2.21}$$

Cette expression montre bien que la longueur $x_\star$ caractérisant la distribution spatiale de la population résulte des actions antagonistes de la diffusion et de la croissance de la population. De même, en utilisant le premier produit adimensionnel, on vérifie que le temps caractéristique $t_\star$ de la population est donné par

$$t_\star \propto \frac{x_\star^2}{\kappa} \propto \frac{1}{\mu} \tag{2.22}$$

L'analyse dimensionnelle ne permet pas d'aller plus loin. On ne peut, par exemple, déterminer les constantes de proportionnalité dans les relations (2.21) et (2.22). La résolution complète de (2.17) fait apparaître en fait des longueur et temps caractéristiques donnés par

$$L_c = \pi\sqrt{\frac{\kappa}{\mu}}, \qquad t_c = \frac{L_c^2}{8\pi^2\kappa} \tag{2.23}$$

Ces expressions sont bien du type fourni par l'analyse dimensionnelle. ◇

La comparaison de nombres adimensionnels obtenus en combinant des grandeurs caractéristiques d'un problème est très souvent utilisée pour comparer les influences respectives de différents processus sur la dynamique d'un système donné.

EXEMPLE 2.8 La composante horizontale de l'équation de la quantité de mouvement d'une particule fluide, s'écrit généralement sous la forme

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{u} + f\,\mathbf{e}_3 \wedge \mathbf{u} = -\nabla_h q + \frac{\partial}{\partial x_3}\left(\tilde{\nu}\frac{\partial \mathbf{u}}{\partial x_3}\right) \tag{2.24}$$

où $\mathbf{u}$ désigne la composante horizontale du vecteur vitesse $\mathbf{v}$, $f = 2\Omega\sin\lambda$ est la fréquence de Coriolis, *i.e.* le double de la composante verticale locale de la vitesse Ω de rotation de la Terre, x_3

est la coordonnée verticale et $\mathbf{e}_3$ le vecteur unitaire correspondant, q est la pression généralisée, $\tilde{\nu}$ représente le coefficient de diffusion turbulente et ∇_h est la composante horizontale de l'opérateur différentiel ∇.

Désignant respectivement par L_c et V_c des longueurs et vitesses caractéristiques de l'écoulement, les ordres de grandeur des termes d'accélération relative et de l'accélération de Coriolis sont donnés respectivement par

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} = O\left(\frac{V_c^2}{L_c}\right) \qquad \text{et} \qquad f\mathbf{E}_z \wedge \mathbf{v} = O\left(fV_c\right) \tag{2.25}$$

L'importance relative des termes d'accélération relative et de Coriolis peut donc être mesurée par le rapport adimensionnel

$$Ro = \frac{\dfrac{V_c^2}{L_c}}{fV_c} = \frac{V_c}{fL_c} \tag{2.26}$$

Ce nombre est appelé le *nombre de Rossby*. Dans le cas d'écoulements aux grandes échelles spatiales et temporelles, on a généralement $Ro \ll 1$, de sorte que l'influence du forçage de Coriolis est prépondérante. ◇

2.5 Détermination systématique des produits adimensionnels.

Dans les exemples précédents, les nombres adimensionnels semblent parfois apparaître miraculeusement. Un méthode systématique existe cependant pour générer les produits adimensionnels. Si on dispose de n nombres dimensionnels $x_1, x_2, \ldots, x_n$ faisant intervenir N dimensions fondamentales D_1, D_2, $\ldots \mathrm{D}_N \in \{\mathrm{M}, \mathrm{L}, \mathrm{T}, \mathrm{I}, \theta, \mathrm{J}\}$ on exprime d'abord les dimensions de chacune des grandeurs x_j, soit

$$[x_j] = \mathrm{D}_1^{a_{1j}} \mathrm{D}_2^{a_{2j}} \ldots \mathrm{D}_N^{a_{Nj}} \tag{2.27}$$

En exprimant que le produit

$$\Pi = x_1^{\alpha_1}\, x_2^{\alpha_2} \cdots x_n^{\alpha_n} \tag{2.28}$$

est sans dimensions, les exposants $\alpha_1, \alpha_2, \ldots, \alpha_n$ apparaissent comme les solutions du système d'équations algébriques linéaires

$$\begin{aligned} \alpha_1 a_{11} + \alpha_2 a_{12} + \cdots + \alpha_n a_{1n} &= 0 \\ \alpha_1 a_{21} + \alpha_2 a_{22} + \cdots + \alpha_n a_{2n} &= 0 \\ &\vdots \\ \alpha_1 a_{N1} + \alpha_2 a_{N2} + \cdots + \alpha_n a_{Nn} &= 0 \end{aligned} \tag{2.29}$$

Ce système comportant N équations pour n inconnues possède au plus $N - n$ solutions linéairement indépendantes qui définissent autant de nombres adimensionnels indépendants.

EXEMPLE 2.9 Le métabolisme d'un poisson peut être décrit par une relation du type

$$T = \alpha W^{\gamma} \tag{2.30}$$

où T désigne le taux de consommation d'oxygène, α les dépenses métaboliques par unité de temps, W la masse du poisson et γ un coefficient approprié.

D'autre part, la croissance est fonction de la ration alimentaire R selon une loi du type

$$\frac{dW}{dt} = R\,\mathrm{e}^{-(a+bR)} \tag{2.31}$$

qui montre que le taux d'assimilation est une fonction décroissante de la ration alimentaire lorsque celle-ci est très grande. Par dérivation, on vérifie aisément que le taux de croissance est maximum pour $R = 1/b$.

Puisque le bilan énergétique s'écrit également

$$\frac{dW}{dt} = R - T \tag{2.32}$$

on en déduit que T est donné par

$$T = R - \frac{dW}{dt} = R(1 - \mathrm{e}^{-(a+bR)}) \tag{2.33}$$

et donc

$$R(1 - \mathrm{e}^{-(a+bR)}) = \alpha W^{\gamma} \tag{2.34}$$

Cette dernière équation peut être utilisée pour calculer, pour un poisson de masse W quelconque, la ration alimentaire nécessaire pour maintenir son activité métabolique.

Le problème ci-dessus fait apparaître les 8 variables t, W, R, T, α, b, a et γ dont les 2 dernières sont déjà sous forme adimensionnelle (comme exposant ou argument d'une fonction transcendante). Pour former des produits adimensionnels à partir des autres variables, déterminons-en d'abord les dimensions. On a

$$\begin{aligned} [t] = \mathrm{T}, \quad [W] = \mathrm{M}, \quad [R] = \mathrm{M}T^{-1}, \\ [T] = \mathrm{M}T^{-1}, \quad [\alpha] = \mathrm{M}^{1-\gamma}\mathrm{T}^{-1}, \quad [b] = \mathrm{M}^{-1}\mathrm{T} \end{aligned} \tag{2.35}$$

où les dimensions de α sont obtenues à partir de l'équations aux dimensions correspondant à (2.30). On remarque que toutes les variables dépendent de deux dimensions fondamentales. Selon le théorème Pi, il leur correspond donc $6-2=4$ produits adimensionnels.

Les produits adimensionnels Π sont obtenus en calculant les produits des variables dimensionnelles, soit

$$\Pi = t^{x_1}\, W^{x_2}\, R^{x_3}\, T^{x_4}\, \alpha^{x_5}\, b^{x_6} \tag{2.36}$$

et en ajustant les exposants x_1, x_2, …, x_6 pour que Π soit adimensionnel. En égalant à 1 les dimensions des deux membres de cette équation, on obtient

$$\begin{aligned} 1 = [\Pi] &= \mathrm{T}^{x_1}\, \mathrm{M}^{x_2}\, (\mathrm{M}T^{-1})^{x_3}\, (\mathrm{M}T^{-1})^{x_4}\, (\mathrm{M}^{1-\gamma}\mathrm{T}^{-1})^{x_5}\, (\mathrm{M}^{-1}\mathrm{T})^{x_6} \\ &= \mathrm{M}^{x_2+x_3+x_4+(1-\gamma)x_5-x_6}\, \mathrm{T}^{x_1-x_3-x_4-x_5+x_6} \end{aligned} \tag{2.37}$$

On en déduit que toute solution $x_1, x_2, \ldots, x_6$ du système d'équations

$$\begin{cases} x_2 + x_3 + x_4 + (1-\gamma)x_5 - x_6 = 0 \\ x_1 - x_3 - x_4 - x_5 + x_6 = 0 \end{cases} \tag{2.38}$$

fournit un nombre adimensionnel. Les équations (2.38) forment un système homogène de 2 équations linéairement indépendantes pour 6 inconnues. Elles possèdent donc une infinité de solutions. Toutes ces solutions s'expriment cependant comme des combinaisons linéaires de 4 (nombre d'inconnues - nombre d'équations linéairement indépendantes) solutions de base. De même, on peut former une infinité de produits adimensionnels (2.36) qui peuvent tous s'écrire comme les produits de certaines puissances de nombres adimensionnels Π_1, Π_2, Π_3, Π_4 de base.

Plutôt que d'appliquer des techniques systématique de résolution de systèmes linéaires (*e.g.* réduction à une forme échelonnée), on préfère faire apparaître chacune des variables les plus significatives dans un et un seul produit adimensionnel qui peut alors être interprété comme l'équivalent adimensionnel de cette variable[4]. Dans le cas présent, on pourra par exemple, choisir d'isoler t, W, R et T pour définir Π_1, Π_2, Π_3 et Π_4

Dans le cas de t, posant $x_1 = 1$, $x_2 = x_3 = x_4 = 0$, le système (2.38) se réduit à

$$\begin{cases} (1-\gamma)x_5 - x_6 = 0 \\ -x_5 + x_6 = -1 \end{cases} \tag{2.39}$$

dont l'unique solution est

$$x_5 = \frac{1}{\gamma}, \qquad x_6 = \frac{1-\gamma}{\gamma} \tag{2.40}$$

Le produit adimensionnel correspondant est donc

$$\Pi_1 = \frac{(\alpha b)^{1/\gamma}}{b} t \tag{2.41}$$

Dans le cas de W, on choisit $x_2 = 1$, $x_1 = x_3 = x_4 = 0$ et on doit résoudre le système

$$\begin{cases} (1-\gamma)x_5 - x_6 = -1 \\ -x_5 + x_6 = 0 \end{cases} \tag{2.42}$$

dont la solution unique est

$$x_5 = x_6 = \frac{1}{\gamma} \tag{2.43}$$

On définit donc

$$\Pi_2 = (\alpha b)^{1/\gamma} W \tag{2.44}$$

De même, on trouve aisément

$$\Pi_3 = b\, R, \qquad \Pi_4 = b\, T \tag{2.45}$$

Les résultats de la mise sous forme adimensionnelle peuvent ensuite être utilisés pour mener des études comparatives sur l'ingestion et le taux de croissance de différents poissons. Les produits

[4] Il peut cependant être impossible de procéder de la sorte pour certaines variables.

adimensionnels montrent que les quantités $b/(\alpha b)^{1/\gamma}$, $1/(\alpha b)^{1/\gamma}$ et $1/b$ représentent des grandeurs caractéristiques qui peuvent être utilisées pour une mise à échelle – respectivement du temps, de la masse et des rations alimentaires et taux de respiration – de variables couvrant différentes échelles de grandeurs.

◇

Chapitre 3

Interpolation.

Tout expérimentateur se retrouve à un moment ou un autre de son analyse face à une série de données correspondant à un certain échantillonnage de grandeurs qui varient de façon continue dans l'espace et/ou dans le temps. Que ce soit pour réaliser l'analyse de ces données, pour les représenter graphiquement ou pour forcer un modèle numérique, il est alors souvent nécessaire de les interpoler pour en reconstituer les variations continues.

L'interpolation de données est une opération a priori anodyne et qui est transparente pour l'utilisateur de beaucoup de logiciels (Excel, Surfer, Matlab...). Il importe cependant d'en connaître les principes pour une bonne utilisation de ces outils et pour éviter les pièges qu'une utilisation irréfléchie peut amener. Nous examinerons successivement les problèmes d'interpolation unidimensionnelle, typiquement des séries temporelles, et les problèmes multidimensionnels comme ceux posés par la représentation de données variables dans l'espace. Enfin, nous traiterons le cas particulier de l'interpolation de données '4D', *i.e.* qui varient à la fois dans l'espace et dans le temps.

3.1 Interpolation unidimensionnelle.

3.1.1 Interpolation linéaire.

La méthode d'interpolation la plus simple et la plus utilisée est l'interpolation linéaire. Celle-ci consiste simplement à faire passer un segment de droite entre deux points de mesure. Désignant par y_1 et y_2 les mesures d'une même variable aux temps t_1 et t_2 ($t_1 < t_2$), il vient simplement

$$y(t) = y_1 + \frac{y_2 - y_1}{t_2 - t_1}(t - t_1) \tag{3.1}$$

Cette formule peut être utilisée pour approcher la variable en tous les instants t compris entre t_1 et t_2. Elle peut également être utilisée avec précaution pour estimer la valeur de y en-dehors de l'intervalle $[t_1, t_2]$. On parle alors d'*extrapolation*. Il convient cependant d'utiliser l'extrapolation avec précaution puisqu'elle revient à étendre l'information fournies par les données en-dehors du domaine d'observation.

Si on dispose de points de mesure (t_k, y_k) $(k = 1, 2 \ldots, N)$ pour une série de N stations temporelles successives (régulièrement espacées dans le temps ou non), on peut construire une interpolation linéaire par morceau en appliquant la formule d'interpolation linéaire (3.1) successivement aux paires de points de mesure successifs $\{(t_k, y_k), (t_{k+1}, y_{k+1})\}$ $(k = 1, 2 \ldots, N-1)$.

Si la dérivée par rapport à t de la grandeur étudiée a un sens, il faut être conscient du fait que l'interpolation linéaire par morceau revient à supposer que cette dérivée est constante par morceau, *i.e.*

$$\frac{y_{k+1} - y_k}{t_{k+1} - t_k} \tag{3.2}$$

et présente donc des discontinuités en chacun des points de support (t_k, y_k) de l'interpolation.

D'autre part, l'interpolation linéaire par morceau construit une représentation continue des données dont la moyenne est en générale différente de celle des données qui ont servi à la construire. Ainsi, par exemple, si les points de support (t_k, y_k) de l'interpolation représentent des flux moyens par mois d'une certaine substance, les bilans mensuels et annuels de cette substance seront biaisés par l'interpolation linéaire par morceau. Si le bilan doit absolument être respecté, la solution la plus simple consiste à approcher l'évolution des flux par des valeurs y_k constantes mois par mois.

3.1.2 Interpolation polynomiale.

L'interpolation polynomiale constitue une extension de l'interpolation linéaire. En effet, l'interpolation linéaire repose sur le fait que par deux points passe une et une seule droite. De même, par trois points passe un et un seul polynôme du second degré. Par quatre points, on peut mener un polynôme unique de degré trois... En général, pour interpoler les données caractérisées par N points de support, on pourra donc utiliser un polynôme de degré $N-1$.

La formule d'interpolation polynomiale de Lagrange permet de construire le polynôme recherché de façon systématique :

$$\begin{aligned} y(t) = y_1 \frac{(t-t_2)(t-t_3)\cdots(t-t_N)}{(t_1-t_2)(t_1-t_3)\cdots(t_1-t_N)} + y_2 \frac{(t-t_1)(t-t_3)\cdots(t-t_N)}{(t_2-t_1)(t_2-t_3)\cdots(t_2-t_N)} \\ + y_N \frac{(t-t_1)(t-t_2)\cdots(t-t_{N-1})}{(t_N-t_1)(t_N-t_2)\cdots(t_N-t_{N-1})} \end{aligned} \tag{3.3}$$

ou, de façon compacte,

$$y(t) = \sum_{i=1}^{N} \left[y_i \left(\prod_{k=1, k\neq i}^{N} \frac{t-t_k}{t_i-t_k} \right) \right] \tag{3.4}$$

L'interpolation polynomiale permet d'obtenir une représentation continûment dérivable des données. Elle permet donc une représentation graphique élégante ainsi que l'estimation des dérivées en chaque point.

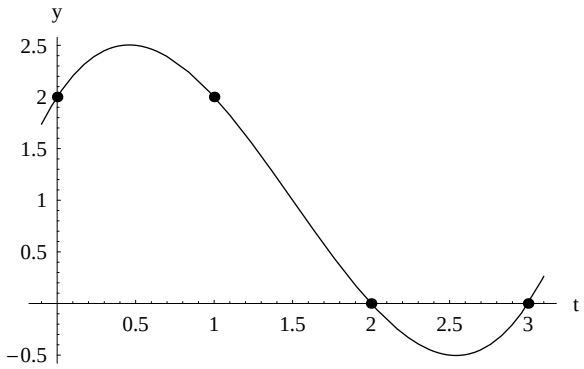


FIG. 3.1 – Interpolation polynomiale des données $\{(0,2),(1,2),(2,0),(3,0)\}$.

Contrairement à l'interpolation linéaire dont les valeurs interpolées sont toujours comprises entre le minimum y_{min} et le maximum y_{max} des données initiales, l'interpolation polynomiale génère souvent des valeurs sortant de l'intervalle $[y_{min}, y_{max}]$. Ainsi, appliquons (3.3) pour interpoler les données $\{(0,2),(1,2),(2,0),(3,0)\}$. Il vient

$$\begin{aligned} y(t) &= 2\frac{(t-1)(t-2)(t-3)}{(0-1)(0-2)(0-3)} + 2\frac{(t-0)(t-1)(t-3)}{(1-0)(1-2)(1-3)} + 0 + 0 \\ &= \frac{2}{3}t^3 - 3t^2 + \frac{7}{3}t + 2 \end{aligned} \tag{3.5}$$

Cette cubique est représentée à la figure 3.1. On constate que le polynôme prend des valeurs supérieures à 2 et inférieures à 0. Si les données à interpoler sont des concentrations, on voit que l'interpolation n'a aucun sens entre $(3,0)$ et $(3,0)$ puisqu'elle y présente des valeurs négatives.

Les problèmes de la figure 3.1 ne font qu'empirer si on augmente l'ordre de l'interpolation : le polynôme risque de présenter un comportement fortement oscillatoire qui n'est absolument pas admissible par rapport à la nature des variables et du problème traités (Cf. figure 3.2).

3.1.3 Interpolation spline.

L'interpolation polynomiale fournit une représentation lisse des données. Cependant, comme le montrent les exemples de la section précédente, elle ne peut être appliquée à un ensemble de données trop important sous peine de donner naissance à des oscillations catastrophique. L'idée de l'interpolation spline est d'éviter ces oscillations en généralisant le concept d'approximation par morceau introduit pour l'interpolation linéaire. Cette fois, la fonction utilisée pour interpoler les données entre deux points de support sera un polynôme de degré p plutôt qu'une expression linéaire.

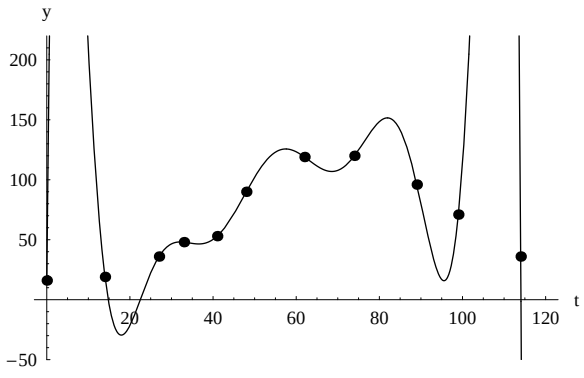


FIG. 3.2 – Interpolation polynomiale de degré élevé.

Considérons les points de support (t_k, y_k) $(k = 1, 2 \ldots, N)$. Une *fonction spline de degré p* est une fonction polynomiale par morceau $y(t)$ telle que

i. dans chaque intervalle $[t_k, t_{k+1}]$, $y(t)$ est un polynôme de degré inférieur ou égal à p ;

ii. $y(t)$ passe par chacun des points de support ;

iii. en chacun des points intérieurs $t_2, t_3, \ldots, t_{N-1}$, $y(t)$ est continue ainsi que ses dérivées jusqu'à l'ordre $p-1$.

Les fonctions spline possèdent de très bonnes propriétés de convergence et de stabilité par rapport aux erreurs d'arrondi. Elles permettent également de très bonnes estimations des dérivées de la fonction interpolée.

La fonction spline cubique ($p = 3$) est la plus utilisée. Examinons en détail comment la construire. En vertu de la première condition, $y(t)$ se réduit à un polynôme d'ordre trois $f_k(t)$ entre deux points de supports consécutifs, i.e.

$$y(t) = f_k(t) = \alpha_k + \beta_k t + \gamma_k t^2 + \delta_k t^3, \qquad t_k \leq t \leq t_{k+1} \qquad (k = 1, 2, \ldots, N-1) \quad (3.6)$$

La fonction spline est donc entièrement définie par la donnée des $4(N-1)$ coefficients α_k, β_k, γ_k et δ_k $(k = 1, 2, \ldots, N-1)$. Les conditions d'interpolation et de continuité de la fonction spline déterminent les conditions que doivent remplir ces coefficients.

– Interpolation :

$$f_k(t_k) = y_k, \qquad k = 1, 2, \ldots, N-1 \quad (3.7)$$

$$f_k(t_{k+1}) = y_{k+1}, \qquad k = 1, 2, \ldots, N-1 \quad (3.8)$$

– Continuité de la dérivée première :

$$f'_{k-1}(t_k) = f'_k(t_k), \qquad k = 2, \ldots, N-1 \quad (3.9)$$

– Continuité de la dérivée seconde :

$$f''_{k-1}(t_k) = f''_k(t_k), \qquad k = 2, \ldots, N-1 \tag{3.10}$$

Les contraintes (3.7)-(3.10) représentent un système de $4N-6$ équations linéaires pour les $4N-4$ inconnues α_k, β_k, γ_k et δ_k. Pour déterminer complètement la fonction spline $y(t)$, on doit donc introduire deux conditions supplémentaires. En général, ces deux conditions supplémentaires prennent l'une des trois formes suivantes :

i. spline naturelle :

$$f''_1(t_1) = f''_{N-1}(t_N) = 0 \tag{3.11}$$

ii. spline périodique :

$$f'_1(t_1) = f'_{N-1}(t_N), \qquad f''_1(t_1) = f''_{N-1}(t_N) \tag{3.12}$$

iii. pentes terminales fixées :

$$f'_1(t_1) = m_1, \qquad f'_{N-1}(t_N) = m_N \tag{3.13}$$

où m_1 et m_N sont des constantes fixées a priori.

La spline cubique naturelle est ainsi nommée car elle rend minimale (dans un certain espace) l'intégrale

$$\int_{t_1}^{t_N} \left[y''(t)\right]^2 dt \tag{3.14}$$

qui constitue une mesure approchée de la courbure totale de la fonction $y(t)$ sur l'intervalle considéré. Vu sous cet angle, la fonction spline naturelle est la fonction la plus régulière d'interpolation des points de support.

À titre d'exemple, la figure 3.3 présente le résultat de l'interpolation des données de la figure 3.2 par une spline cubique naturelle. Le résultat est bien plus satisfaisant que celui obtenu par l'interpolation polynomiale.

La détermination complète des coefficients α_k, β_k, γ_k et δ_k définissant la fonction spline cubique passe par la résolution d'un système linéaire d'équations liant tous les points de support. C'est donc une approximation globale de la fonction ; si on ajoute de nouveaux points ou si on modifie un point de support, tous les polynômes cubiques définissant $y(t)$ sont modifiés (et toute la procédure de calcul doit être répétée). Cependant, en raison de la structure particulière du système d'équations linéaires correspondant à (3.7)-(3.10) et aux conditions terminales, l'effet d'une modification d'un point de support s'atténue rapidement lorsque la distance au point de support perturbé augmente.

On remarque sur la figure 3.3 que l'interpolation spline n'induit qu'un très faible dépassement de la valeur maximale contenue dans les données. Un tel dépassement (ou une valeur inférieure au minimum des données) est toujours possible et est en général tout à faire réaliste ; il serait tout à fait fortuit et surprenant que l'échantillonnage des données capte parfaitement les extrema. Cependant, il est également possible que cette valeur ne puissent exister dans la nature. De nombreuses variantes de fonctions

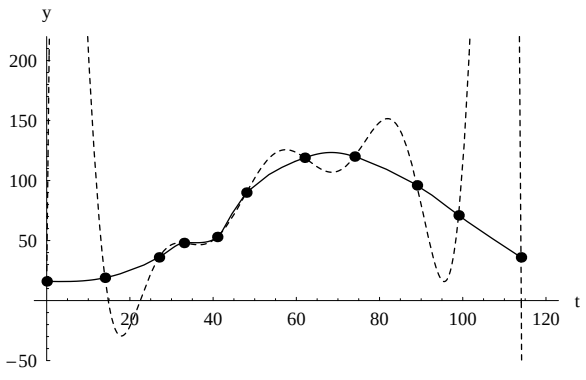


FIG. 3.3 – Interpolation spline cubique nature (trait continu) et polynomiale de degré élevé (trait interrompu).

spline sont alors disponibles pour modifier l'interpolation ponctuellement et éviter ce problème. L'interpolation spline sous tension, par exemple, repose sur le même principe d'approximation par morceau mais remplace les polynômes de l'interpolation spline classique par des combinaisons de polynômes et de fonctions exponentielles. Ces combinaisons dépendent d'un paramètre permettant de contrôler localement 'la tension' et d'éliminer les points d'inflexion jugés superflus.

3.2 Interpolation multi-dimensionnelle.

Dans la section précédente, les données interpolées dépendaient a priori d'une seule variable indépendante, notée t. Penchons nous maintenant sur l'interpolation de grandeurs qui dépendent a priori de plusieurs variables indépendantes. Bien que la plupart des concepts introduits dans la suite peuvent être aisément généralisés à des problèmes en dimension quelconque, nous aurons à l'esprit le problème de l'interpolation spatiale bi-dimensionnelle.

Les techniques sont couramment employées pour ramener les données d'une grille expérimentale irrégulière vers une grille régulière en vue de leur analyse ou de leur représentation graphique.

3.2.1 Interpolation bi-linéaire.

En présence de données localisées de façon quelconque dans un plan, on peut aisément généraliser l'interpolation linéaire en deux étapes.

Tout d'abord, on découpe le plan en sous-domaines triangulaires dont les sommets sont les points de mesure.

En général, on peut découper le domaine étudié de différentes façons. Dans la

triangulation de Delaunay, on choisit de former un triangle à partir de trois points si et seulement si le cercle circonscrit à ce triangle ne contient aucun autre point de mesure.

Ensuite, dans chaque triangle, on approche la variable étudiée par une fonction linéaire

$$f(x,y) = \alpha + \beta x + \gamma y \tag{3.15}$$

Si la variable prend les valeurs z_1, z_2 et z_3 aux sommets (x_1,y_1), (x_2,y_2), (x_3,y_3) du triangle, alors f_k représente une interpolation des points de support correspondant si

$$f(x_i,y_i) = z_i \qquad i = 1,2,3 \tag{3.16}$$

Ce système de trois équations à 3 inconnues possède une solution unique (sauf si les points (x_1,y_1), (x_2,y_2), (x_3,y_3) sont alignés auquel cas le triangle dégénère en un segment de droite) qui détermine de façon unique l'interpolation linéaire à l'intérieur du triangle considéré.

En répétant l'opération séparément dans chacun des triangles, on construit une interpolation linéaire par morceau. Celle-ci est continue d'un triangle à l'autre. Le long d'une arête formant la frontière entre deux triangles, l'interpolation bi-linéaire (3.15) se réduit à une interpolation linéaire entre les valeurs prises par le champs aux deux sommets d'extrémité de cette arête. Deux triangles contigus présentant une arête commune partagent également les sommets correspondant et les interpolations bi-linéaires associées coïncident donc sur cette arête. Les dérivées partielles, constantes sur chaque triangle, sont par contre discontinues d'un triangle à l'autre.

Remarquons que la procédure permet d'interpoler les données dans l'union des triangles, soit l'enveloppe convexe des points de mesure. Toute tentative de détermination du champ en dehors de cette enveloppe constitue une dangereuse extrapolation.

3.2.2 Interpolation par distance inverse.

Bien que l'on puisse généraliser à plusieurs dimensions les interpolations polynomiale et spline introduites dans le cadre unidimensionnelle, l'interpolation par distance inverse est souvent préférée à ces techniques généralisées.

Le principe de la méthode est de calculer la valeur du champ en chaque point à partir d'une moyenne pondérée des mesures disponibles. Pour que les données proches du point étudié interviennent davantage dans la moyenne que les données plus éloignées, les poids sont inversement proportionnels à une certaine puissance $p > 0$ de la distance entre le point courant et le point de mesure. En présence de données (x_i,y_i,z_i) $(i = 1,2,\ldots,N)$, où z_i désigne les valeurs prises par la grandeur à interpoler, on aura donc

$$z(x,y) = \sum_{i=1}^{N} w_i(x,y) z_i \tag{3.17}$$

où les poids w_i sont donnés par

$$w_i(x,y) = \frac{h_i^{-p}(x,y)}{\sum_{j=1}^{N} h_j^{-p}(x,y)} \qquad \text{où} \qquad h_i(x,y) = \sqrt{(x-x_i)^2 + (y-y_i)^2} \tag{3.18}$$

Par construction, les valeurs du champ interpolé seront comprises entre les valeurs maximales et minimales des données initiales. L'interpolation est également indéfiniment continûment dérivable.

L'exposant p dans (3.17) est souvent pris égal à 2. Plus l'exposant p est élevé, moins les points éloignés influencent la valeur locale de l'interpolation. Pour éviter l'influence des points trop éloignés, on peut aussi ne prendre en compte dans (3.17) que les points dont la distance au point courant d'interpolation est inférieure à une certaine distance limite (calculée pour avoir suffisamment de points de mesure ou fixée à partir de caractéristiques connues du champ étudié). De façon alternative, on peut également restreindre les données intervenant dans (3.17) aux seuls données correspondant aux sommets d'un triangle de Delaunay.

3.3 Estimation linéaire.

Les techniques exposées ci-dessus sont toutes de vraies méthodes d'interpolations : le champ interpolé passe exactement par les points de support. Ceci n'est cependant pas toujours souhaitable si on prend en compte, par exemple, l'erreur expérimentale associée à chacune des mesures ou la variabilité naturelle de la grandeur observée. Il arrive par exemple fréquemment que des valeurs très différentes soient obtenues pour des mesures effectuées en des points très proches. Si une procédure d'interpolation vraie, comme celles des sections précédentes, est appliquée à ces données, des gradients spatiaux peu réalistes apparaîtront. Dans ce cas, il faut se résoudre à relaxer la contrainte d'interpolation exacte des mesures. On construit alors une approximation qui approche 'au mieux' les données disponibles dans un certain sens à définir.

Dans cette section, nous examinons les problèmes d'estimation linéaire, *i.e.* ceux dans lesquels les paramètres de l'interpolation apparaissent linéairement.

3.3.1 Problème de base de régression linéaire.

Le problème de base de la régression linéaire consiste à estimer, à partir des données expérimentales (x_i, y_i) $(i = 1, 2 \ldots, N)$, les paramètres b_0 et b_1 d'une loi liant les valeurs d'une variable aléatoire y à une variable indépendante non-aléatoire x selon le modèle

$$y = E[y] + \varepsilon = b_0 + b_1 x + \varepsilon \tag{3.19}$$

où $E[y]$ désigne l'espérance mathématique de la variable aléatoire y et ε est une variable aléatoire de moyenne nulle. En d'autres termes, on suppose que, pour chaque valeur de la variable indépendante x, il existe une distribution aléatoire de y dont la valeur mesurée constitue une réalisation. La moyenne de la population correspondant à la variable indépendante x est donnée par la loi linéaire $b_0 + b_1 x$. Le paramètre b_1 mesurant la sensibilité de $E[y]$ à x est appelé le *coefficient de régression.*

L'application des techniques de régression linéaire aux données expérimentales suppose que l'écart entre la prévision $\hat{y}_i = b_0 + b_1 x_i$ du modèle linéaire au point x_i et

l'observation y_i en ce même point reflète le caractère aléatoire de la variable y. Le terme aléatoire ε est supposé représenter l'erreur de mesure et la variabilité naturelle superposée au modèle linéaire $b_0 + b_1 x$.

Pour déterminer les coefficients b_0 et b_1, on minimise la somme des carrés des écarts SSE entre les valeurs prédites $\hat{y}_i$ et réellement observées y_i, soit

$$SSE = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} \Big[y_i - (b_0 + b_1 x_i) \Big]^2 \tag{3.20}$$

L'annulation des dérivées partielles de (3.20) par rapport à b_0 et b_1, fournit les équations linéaires permettant de déterminer les coefficients qui rendent minimale la somme des carrés des écarts. On a

$$\begin{cases} \dfrac{\partial SSE}{\partial b_0} = -2 \displaystyle\sum_{i=1}^{N} (y_i - b_0 - b_1 x_i) = 0 \\ \dfrac{\partial SSE}{\partial b_1} = -2 \displaystyle\sum_{i=1}^{N} x_i (y_i - b_0 - b_1 x_i) = 0 \end{cases} \tag{3.21}$$

On calcule aisément

$$b_1 = \frac{N \displaystyle\sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i \right) \left(\sum_{i=1}^{N} y_i \right)}{N \displaystyle\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} = \frac{\displaystyle\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\displaystyle\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{3.22}$$

où on a noté

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \qquad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{3.23}$$

À partir de (3.22), on en déduit alors que

$$b_0 = \bar{y} - b_1 \bar{x} \tag{3.24}$$

Cette dernière équation montre que la droite de régression passe toujours par le point moyen $(\bar{x}, \bar{y})$. La droite de régression partage en fait les points expérimentaux en deux groupes dont les écarts par rapport à la droite de régression sont respectivement positifs et négatifs et se compensent exactement.

Introduisant les notations

$$SST = \sum_{i=1}^{N} (y_i - \bar{y})^2, \qquad SSR = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 \tag{3.25}$$

et utilisant (3.22)-(3.24) on peut écrire (3.20) sous la forme

$$SSE = SST - SSR \tag{3.26}$$

Dans cette expression, SST représente la variance totale de la variable dépendante y, SSR mesure la variance expliquée par le modèle de régression tandis que SSE représente la variance totale qui n'est pas expliquée (ou représentée) par le modèle de régression linéaire.

Clairement, le rapport

$$r^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{3.27}$$

entre la variance expliquée par le modèle linéaire et la variance totale des données constitue une mesure de la validité du modèle linéaire de régression. Il est appelé le *coefficient de détermination*. Par construction, celui-ci varie entre 0 et 1. Une valeur proche de 1 indique un bon accord entre les données et le modèle linéaire. Lorsque l'accord est de moins en moins bon, r^2 décroît vers sa valeur minimale possible de zéro.

Le coefficient de détermination peut être rapproché du *coefficient de corrélation r* qui peut être défini pour deux variables aléatoires[1] x et y par

$$r = \frac{C_{xy}}{s_x \, s_y} \tag{3.28}$$

où

$$C_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \tag{3.29}$$

désigne la *covariance* de x et y et où

$$s_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}, \qquad s_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{3.30}$$

sont les estimateurs des écart-types de deux variables aléatoires. Comme le suggère la notation, le coefficient de détermination est le carré du coefficient de corrélation. À ce titre, il est également adimensionnel et il varie entre -1 et +1. Un coefficient de corrélation positif signifie que les variables x et y varient en phase, *i.e.* dans le même sens. Un coefficient négatif témoigne de variations en opposition de phase, *i.e.* à une augmentation d'une variable correspond une diminution de l'autre variable.

Il faut cependant se garder d'un optimisme excessif face à un coefficient de détermination proche de l'unité. En effet, la mesure absolue de l'accord entre les données et le modèle linéaire est donnée par l'*erreur standard de l'estimation*

$$s_\varepsilon = \sqrt{\frac{SSE}{N-2}} = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{3.31}$$

[1] En introduisant le problème de régression linéaire, nous avons supposé que seule la variable y était aléatoire.

Le facteur $N-2$ intervenant au dénominateur de (3.31) représente une estimation du nombre de degrés de liberté de ε basée sur le fait que les deux paramètres b_0 et b_1 sont estimés à partir de l'ensemble des données. Dans le cas limite d'un ensemble de 2 points de mesure, la régression linéaire conduit à interpoler parfaitement les données aux points de mesure et le coefficient de détermination est unitaire. Cependant, l'erreur standard de l'estimation est théoriquement infinie ; on ne dispose pas de suffisamment d'information complémentaire pour quantifier l'erreur associée à la régression linéaire. Des statistiques précises sur les erreurs associées au modèle linéaire ne peuvent être obtenues que si on augmente le nombre de points de mesure.

Pour aller plus loin dans l'analyse de l'erreur, il est nécessaire de faire des hypothèses sur le type de distribution statistique de l'erreur ε. Si on suppose que la distribution de celle-ci est normale (moyenne nulle) et indépendante de la variable aléatoire x, alors l'erreur standard de l'estimation peut être utilisée pour construire un intervalle de confiance pour l'estimation. Ainsi, à peu près 68.3% des observations se situeront dans un intervalle de $\pm 1 s_\varepsilon$ de la droite de régression, 95.4% tomberont à $\pm 2 s_\varepsilon$ et l'intervalle $\pm 3 s_\varepsilon$ contiendra 99.7% des données.

D'autres option sont disponibles dans les logiciels de traitement statistique. Par exemple, sous l'hypothèse de normalité de ε utilisée plus haut et considérant le coefficient de régression comme une variable aléatoire, on peut tester l'hypothèse nulle $b_1 = 0$, *i.e.* pas de corrélation entre les deux variables x et y. Lorsque l'hypothèse nulle est rejetée (pour un niveau de confiance α donné), la corrélation est déclarée statistiquement significative.

Remarquons qu'un coefficient de corrélation élevé ou un bon accord d'une droite de régression $y(x)$ avec les données expérimentales ne signifie pas que x est la cause de y. Il se peut très bien que y soit la cause de x ou que x et y soient influencés par un même facteur (ou une combinaison de facteurs). Cette dernière possibilité est très utilisée dans les études climatiques et dans certaines études environnementales dont on ne parvient pas vraiment à cerner ou à définir les paramètres clés. On définit dans ce cas un indicateur ou 'proxy' comme une variable aisément mesurable et caractéristique du changement climatique ou environnemental.

3.3.2 Estimation au sens des moindres carrés.

Le problème (3.19) est qualifié de linéaire, non pas parce que la partie déterministe varie linéairement avec la variable indépendante x mais parce que les paramètres inconnus b_0 et b_1 y apparaissent linéairement. C'est cette propriété qui permet d'obtenir des équations linéaires lorsque l'on minimise les écarts au sens des moindres carrés. La méthode est donc généralisable à d'autres modèles que le modèle (3.19) pourvu que les paramètres inconnus y interviennent linéairement.

Parmi les modèles les plus couramment utilisés, citons

– le modèle logarithmique

$$y(x) = b_0 + b_1 \ln x, \tag{3.32}$$

– le modèle polynomial

$$y(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_k x^k \tag{3.33}$$

D'autres modèles, a priori non linéaires en les paramètres, peuvent être traités après transformation des données. Ainsi, le modèle

$$y(x) = \alpha \, e^{\beta x} \tag{3.34}$$

peut être ramené à un problème de régression linéaire des données modifiées $(\tilde{x}_i, \tilde{y}_i) = (x_i, \ln y_i)$ puisque, en prenant le logarithme des deux membres de (3.34), on obtient

$$\ln y = \ln \alpha + \beta x, \qquad \text{soit} \qquad \tilde{y} = b_0 + b_1 \tilde{x} \tag{3.35}$$

où $b_0 = \ln \alpha$ et $b_1 = \beta$.

De même, une relation du type

$$y(x) = \alpha x^{\beta} \tag{3.36}$$

constitue une relation linéaire pour les variables $(\tilde{x}, \tilde{y}) = (\ln x, \ln y)$ puisque

$$\ln y = \ln \alpha + \beta \ln x \tag{3.37}$$

La relation

$$y(x) = \frac{\alpha}{\beta + x} \tag{3.38}$$

peut donner lieu à deux types de linéarisation. Soit

$$\frac{1}{y} = \frac{\beta}{\alpha} + \frac{x}{\alpha} \tag{3.39}$$

pour les variables $(\tilde{x}, \tilde{y}) = (x, 1/y)$ ou

$$y = \frac{\alpha}{\beta} + \frac{-1}{\beta} xy \tag{3.40}$$

pour les variables $(\tilde{x}, \tilde{y}) = (xy, y)$.

Enfin les paramètres de la loi de Michaelis-Menten

$$y(x) = \frac{\alpha x}{\beta + x} \tag{3.41}$$

peuvent être estimés par régression linéaire appliquée à la relation transformée

$$\frac{1}{y} = \frac{1}{\alpha} + \frac{\beta}{\alpha} \frac{1}{x} \tag{3.42}$$

pour les variables $(\tilde{x}, \tilde{y}) = (1/x, 1/y)$ ou

$$y = \alpha + (-\beta) \frac{y}{x} \tag{3.43}$$

pour les variables $(\tilde{x},\tilde{y}) = (y/x, y)$.

Remarquons qu'une optimisation directe (au sens des moindres carrés) des paramètres des relations (3.34), (3.36), (3.38), (3.41) est possible en utilisant des algorithmes spécialisés d'optimisation mathématique. Les paramètres optimaux obtenus seront en général légèrement différents de ceux obtenus par régression linéaire. Les différentes linéarisations d'une même relation conduiront également à des résultats différents. En effet, chacune des optimisations correspond à une pondération différente des écarts.

3.4 Analyse objective.

L'ajustement d'une droite de régression à un ensemble de données constitue une alternative à l'interpolation pure et dure tenant compte du fait que les données sont imparfaites et ne doivent donc pas être représentées exactement par le modèle. La même approche peut être étendue à des données distribuées dans l'espace.

Le but de l'analyse objective est en général de représenter sur une grille régulière des données expérimentales distribuées de façon quelconque. L'analyse est dite objective si l'interpolation est guidée par une logique mathématique bien définie. L'interpolation est aussi qualifiée d'optimale si elle correspond, comme dans le cas de la régression linéaire, à la minimisation d'une certaine mesure de l'erreur (en général l'erreur quadratique de l'estimation). L'interpolation optimale se base sur l'hypothèse de stationnarité (indépendance par rapport au temps pendant la période correspondant aux mesures) et d'homogénéité spatiale (indépendance par rapport à l'espace) des caractéristiques statistiques des données étudiées. Grace à ces hypothèses, les statistiques de l'erreur peuvent être estimées à partir des données elle-mêmes et utilisées pour en fournir une représentation continue.

Pour fixer les idées, considérons un ensemble de mesures $d_i = d(\mathbf{x}_i, t_i)$ obtenues à différentes positions $\mathbf{x}_i$ en des temps différents t_i. Le problème de l'analyse objective consiste à recréer une approximation continue $D_a(\mathbf{x}, t)$ du champ réel inconnu $D(\mathbf{x}, t)$ à partir de ces données.

La première étape du traitement consiste généralement en le calcul de la différence entre les observations et un champ de référence D^{ref}, ce qui définit l'*anomalie* du champ étudié. Ceci permet de soustraire des observations la *moyenne* et la *tendance* (pas nécessairement linéaire) connues pour le champ étudié et de se concentrer sur l'interpolation optimale de l'anomalie. L'anomalie possède, en général, des caractéristiques statistiques plus intéressantes que le champ de départ comme la stationnarité, l'homogénéité ou l'isotropie.

Le champ de référence est en général fourni par des données climatiques ou historiques. Lorsque de telles données ne sont pas disponibles on peut également évaluer la moyenne et la tendance directement à partir des données expérimentales que l'on désire interpoler. Le champ de référence constitue une première approximation du champ étudié qui permet de décrire les zones qui ne sont couvertes par aucune donnée, d'introduire des structures connues (zone frontale, upwelling,...) qui ne sont pas bien représentées par les

données ou de forcer une certaine cohérence avec la dynamique connue de la région.

Les méthodes d'interpolation optimales font partie de l'arsenal des méthodes d'estimation linéaire. Le champ D_a est reconstruit sous la forme d'une somme pondérée des différentes mesures disponibles

$$D_a(\mathbf{x},t) = D^{ref}(\mathbf{x},t) + \sum_{i=1}^{N} w_i(\mathbf{x},t)(d_i - D^{ref}(\mathbf{x}_i,t_i)) \tag{3.44}$$

où $w_i(\mathbf{x},t)$ représente le poids de la mesure i dans la reconstruction du champ. Notant les anomalies par d_i' et D_a', cette équation peut s'écrire plus simplement

$$D_a'(\mathbf{x},t) = \sum_{i=1}^{N} w_i(\mathbf{x},t) d_i' \tag{3.45}$$

La relation (3.44) est semblable à celle utilisée dans l'interpolation par distance inverse. Cependant, on désire maintenant ajuster les poids w_i pour refléter, non seulement la proximité des données au point courant, mais également la fiabilité et la précision des données. En effet, l'interpolation optimale ne devant pas passer exactement par les points expérimentaux, il importe de quantifier la contrainte représentée par les mesures.

Idéalement, les poids devront prendre en compte l'erreur associée à chaque type de mesure et varier de façon inversement proportionnel à l'erreur expérimentale. Ainsi, s'il s'agit de reconstituer le champ de température, on pondérera différemment les mesures fournies par XBT, CTD et observations satellitaires. L'erreur de mesure peut également être influencée par le type de traitement préalable subi par ces mesures. Par exemple, la constitution de 'super-observations' permet de construire des nouvelles données entachées d'une erreur minimale en prenant la moyenne de plusieurs mesures réalisées en des points très proches l'un de l'autre.

Les poids w_i doivent également refléter l'écart attendu par rapport au champ de référence, *i.e.* l'ordre de grandeur normale de l'anomalie. Si le champ de référence est constitué d'une moyenne climatique, l'écart-type associé constitue une mesure appropriée de cet écart. Si le champ de référence est construit à partir des résultats d'un modèles numériques de prévision, les statistiques d'erreur du modèle pourront être prise en compte.

Enfin, l'ajustement des poids w_i doit permettre de compenser l'inhomogènéité de la distribution des points de mesure. Pour comprendre la problématique liée à la distribution des mesures, considérons les trois cas particuliers de distribution des points expérimentaux de la figure 3.4. Dans les trois cas, les points de mesure sont situés sur un même cercle centré sur le point courant où on désire estimer la valeur du champ.

i. Dans le premier cas, les données sont réparties aux sommets d'un triangle équilatéral. La valeur au milieu doit donc être également influencée par chacune des mesures (si les erreurs et incertitudes sur les mesures sont égales). En particulier, si les données sont parfaites, on doit logiquement poser

$$w_1 = w_2 = w_3 = \frac{1}{3} \tag{3.46}$$

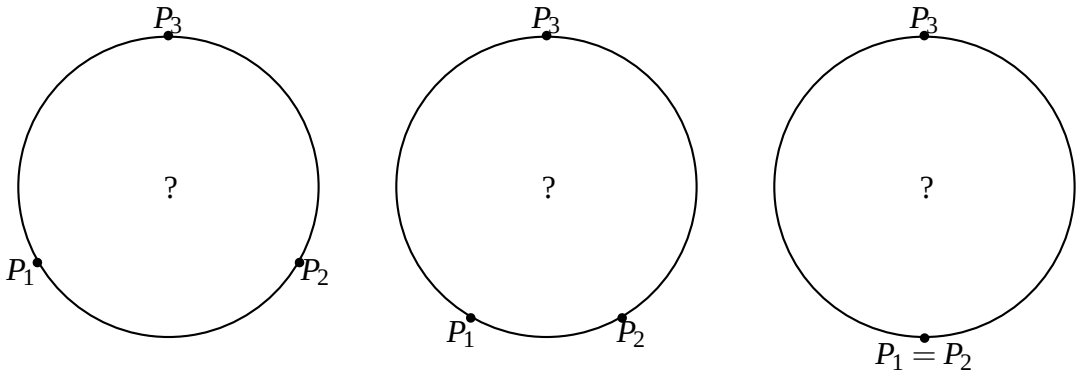
ii. Dans le deuxième cas, les points de mesure 1 et 2 sont proches l'un de l'autre et devraient logiquement apporter des informations proches. La quantité totale d'information apportée par les points 1 et 2 est inférieure à celle apportée dans le premier cas. Si les données sont parfaites on doit donc prendre

$$w_1 = w_2 = \frac{1}{3} - \varepsilon, \quad w_3 = \frac{1}{3} + 2\varepsilon \tag{3.47}$$

pour une certaine valeur de ε.

iii. Dans le cas dégénéré où les points 1 et 2 sont confondus et apportent exactement la même information, on aura

$$w_1 = w_2 = \frac{1}{4}, \quad w_3 = \frac{1}{2} \tag{3.48}$$



i. ii. iii.

FIG. 3.4

Pratiquement, le champs reconstruit D_a est lui même échantillonné sur une grille d'analyse régulière qui induit un nouveau filtrage. Selon le théorème d'échantillonnage de Nyquist, la plus petite longueur décrite par les données est égale à deux fois la distance entre les points d'observations. La taille de la matrice doit être choisie en conséquence.

Considérons désormais que D_a représente la matrice colonne des anomalies (on a laisser tomber les $'$ pour alléger les notations) en les différents points de la grille d'analyse tandis que d représente la matrice colonne des observations. La relation (3.45) peut s'écrire sous forme matricielle selon

$$\mathsf{D}_a = \mathsf{W}\mathsf{d} \tag{3.49}$$

où W désigne la matrice des poids $w_j(x_i, t_i)$. Les éléments de W sont les inconnues du problème.

Le champ réel est donné par

$$\mathsf{D} = \mathsf{D}_a + \varepsilon_a \tag{3.50}$$

où ε_a désigne la matrice colonne des erreurs associées au champ interpolé. De même, désignons par ε_o la matrice colonne des erreurs associées aux mesures. Évidemment, ces grandeurs ne sont pas connues, puisque le champ réel n'est pas connu ; dans la suite, on émettra cependant une série d'hypothèses les concernant et permettant de résoudre le problème. Pour l'instant, on suppose seulement qu'il n'y a pas d'erreur systématique dans les observations, *i.e.* que celles-ci ne sont pas biaisées.

À partir de ces grandeurs, on définit les matrice de covariances des erreurs C_ε obtenue en calculant le produit de la matrice colonne d'erreur ε par sa transposée et en en prenant l'espérance mathématique, *i.e.*

$$\mathsf{C}_\varepsilon = E[\varepsilon_a {\varepsilon_a}^{\mathrm{T}}] = \begin{pmatrix} E[e_1e_1] & E[e_1e_2] & \cdots & E[e_1e_n] \\ E[e_1e_1] & E[e_1e_2] & \cdots & E[e_1e_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[e_ne_1] & E[e_ne_2] & \cdots & E[e_ne_n] \end{pmatrix} \tag{3.51}$$

où $e_i = D_a(x_i,t_i) - D(x_i,t_i)$ désigne l'erreur du champ interpolé au point (x_i,t_i) de la grille d'analyse. Remarquons que la matrice de covariance est symétrique et définie positive. Sur sa diagonale, on trouve les variances $E[e_ie_i] = \sigma_i^2$ des erreurs aux différents points de la grille d'analyse. Les termes non diagonaux décrivent les dépendances entre les différents points de la grille d'analyse.

En utilisant (3.49), la matrice C_ε prend la forme

$$\mathsf{C}_\varepsilon = E[(\mathsf{Wd}-\mathsf{D})(\mathsf{Wd}-\mathsf{D})^{\mathrm{T}}] = E[\mathsf{Wdd}^{\mathrm{T}}\mathsf{W}^{\mathrm{T}} - \mathsf{Dd}^{\mathrm{T}}\mathsf{W}^{\mathrm{T}} - \mathsf{WdD}^{\mathrm{T}} + \mathsf{DD}^{\mathrm{T}}] \tag{3.52}$$

Introduisons les matrices de covariance du champ C_D, des observations C_d et la matrice de covariance conjointe du champ et des observations C_{Dd} telles que

$$\mathsf{C}_D = E[\mathsf{DD}^{\mathrm{T}}], \qquad \mathsf{C}_d = E[\mathsf{dd}^{\mathrm{T}}], \qquad \mathsf{C}_{Dd} = E[\mathsf{Dd}^{\mathrm{T}}] \tag{3.53}$$

Avec ces notations, (3.52) devient

$$\mathsf{C}_\varepsilon = \mathsf{W}\mathsf{C}_d\mathsf{W}^{\mathrm{T}} - \mathsf{C}_{Dd}\mathsf{W}^{\mathrm{T}} - \mathsf{W}{\mathsf{C}_{Dd}}^{\mathrm{T}} + \mathsf{C}_D \tag{3.54}$$

qui fait clairement apparaître la dépendance de C_ε en la matrice des poids W.

La clé de l'analyse objective réside dans le choix de W minimisant les termes diagonaux (ou plus exactement la trace) de la matrice de covariance C_ε de l'erreur d'analyse, *i.e.* les variances des erreurs aux différents points de la grille d'analyse. On montre[2] que ce minimum est atteint pour

$$\mathsf{W} = \mathsf{C}_{Dd}\mathsf{C}_d^{-1} \tag{3.55}$$

[2]En utilisant la symétrie de C_d, on peut écrire C_ε sous la forme

$$\mathsf{C}_\varepsilon = (\mathsf{W}-\mathsf{C}_{Dd}\mathsf{C}_d^{-1})\mathsf{C}_d(\mathsf{W}-\mathsf{C}_{Dd}\mathsf{C}_d^{-1})^{\mathrm{T}} - \mathsf{C}_{Dd}\mathsf{C}_d^{-1}{\mathsf{C}_{Dd}}^{\mathrm{T}} + \mathsf{C}_D$$

En utilisant le caractère symétrique défini positif de C_d, on en déduit que

$$(\mathsf{W}-\mathsf{C}_{Dd}\mathsf{C}_d^{-1})\mathsf{C}_d(\mathsf{W}-\mathsf{C}_{Dd}\mathsf{C}_d^{-1})^{\mathrm{T}} \qquad \text{et} \qquad \mathsf{C}_{Dd}\mathsf{C}_d^{-1}{\mathsf{C}_{Dd}}^{\mathrm{T}}$$

L'estimateur, dit de Gauss-Markov, correspondant au minimum au sens des moindres carrés est donc donné par

$$\mathsf{D}_a = \mathsf{C}_{Dd}\mathsf{C}_d^{-1}\mathsf{d} \tag{3.56}$$

L'erreur associée est quant-à elle caractérisée par

$$\mathsf{C}_\varepsilon = \mathsf{C}_D - \mathsf{C}_{Dd}\mathsf{C}_d^{-1}\mathsf{C}_{Dd}{}^{\mathrm{T}} \tag{3.57}$$

Le calcul de W selon (3.55) suppose que l'on dispose des matrices C_{Dd} et C_d. Pour ce faire, on devrait disposer de séries de mesures appropriées pour pouvoir calculer les espérances mathématiques qui se cachent derrière ces matrices. Il est parfois possible d'utiliser des données historiques ou climatiques, si on peut supposer qu'aucun changement n'est intervenu dans le système entre la période d'origine de ces données et le moment étudié. Dans un grand nombre de cas, cependant, on ne dispose que des mesures d effectuées en un nombre limité de points. L'idée est alors de supposer que les statistiques sont homogènes, stationnaires et isotropes dans la région étudiée. Dans ce cas, les corrélations d'un grandeur mesurée en deux points $\mathbf{x}_i$ et $\mathbf{x}_j$ ne dépendent plus que de la distance $|\mathbf{x}_i - \mathbf{x}_j|$ entre ces points. Dès lors,

$$(\mathsf{C}_d)_{ij} = E[d_i d_j] \approx Cov(|\mathbf{x}_i - \mathbf{x}_j|) \tag{3.58}$$

où le membre de gauche représente la covariance calculée à partir de toutes les paires de mesures distantes (approximativement) de $|\mathbf{x}_i - \mathbf{x}_j|$. Pour être sur que l'estimation de la covariance produise une matrice C_d symétrique définie positive, une fonction suffisamment régulière peut être ajustée aux valeurs calculées à partir de (3.58).

Remarquons que la matrice C_d contient en elle à la fois l'influence des erreurs expérimentales et de la corrélation entre les données vraies. En effet, notant ε_o cette erreur expérimentale et D_o le champ réel aux points de mesure, on a

$$\mathsf{d} = \mathsf{D}_o + \varepsilon_o \tag{3.59}$$

et

$$\mathsf{C}_d = E[\mathsf{d}\mathsf{d}^{\mathrm{T}}] = E[\mathsf{D}_o\mathsf{D}_o{}^{\mathrm{T}} + \mathsf{D}_o\varepsilon_o{}^{\mathrm{T}} + \varepsilon_o\mathsf{D}_o{}^{\mathrm{T}} + \varepsilon_o\varepsilon_o{}^{\mathrm{T}}] \tag{3.60}$$

Si les erreurs expérimentales ne sont pas corrélées avec le champ réel

$$E[\mathsf{D}_o\varepsilon_o{}^{\mathrm{T}}] = E[\varepsilon_o\mathsf{D}_o{}^{\mathrm{T}}] = 0 \tag{3.61}$$

et (3.60) se simplifie selon

$$\mathsf{C}_d = E[\mathsf{D}_o\mathsf{D}_o{}^{\mathrm{T}}] + E[\varepsilon_o\varepsilon_o{}^{\mathrm{T}}] \tag{3.62}$$

possèdent des éléments diagonaux positifs. Les éléments diagonaux de C_ε seront donc minimaux pour

$$\mathsf{W} = \mathsf{C}_{Dd}\mathsf{C}_d^{-1}$$

Une procédure semblable à celle utilisée pour calculer C_d peut être utilisée pour estimer C_{Dd} à partir des données mesurée d.

Remarquons que l'équation (3.55) montre que le poids des observations est inversement proportionnel à la covariance des données C_d. Ceci est bien conforme aux principes généraux énoncés précédemment ; si une variable présente une forte variabilité, son poids dans l'interpolation doit être réduit. De même, si les erreurs expérimentales sont entachées d'une forte erreur expérimentale prenant la forme d'un bruit blanc, la matrice $E[\varepsilon_o {\varepsilon_o}^{\mathrm{T}}]$ est diagonale, les coefficients diagonaux de C_d sont grands et le poids des points incriminés est réduit. Enfin, remarquons que l'erreur (3.57) augmente lorsque l'incertitude et/ou l'erreur sur les mesures augmentent.

3.5 Krigeage

Dans la méthode du krigeage ('kriging' en anglais) on calcule une moyenne pondérée des observations semblables à celle utilisée dans l'interpolation par distance inverse (3.17). Les poids sont cependant choisis différemment. Ils sont déterminés après une étude de la variabilité spatiale des données à représenter.

Tout commence par la construction d'un *semi-variogramme* (généralement appelé abusivement variogramme) montrant les variations de la corrélation entre les données en fonction de la distance d entre celles-ci. Pour construire le variogramme, on groupe toutes les données par paires et on répartit ces couples dans différentes classes en fonction de la distance qui les sépare. Le nombre de classes doit être suffisant pour décrire correctement l'influence de la distance entre données mais doit également être limité pour disposer de suffisamment de couples dans chaque classe de façon à pouvoir en tirer des résultats statistiquement significatifs. On pourra par exemple fixer le nombre de classes Nc en fonction de la règle de Sturge, soit

$$Nc = \lfloor 1 + 3.3 \log_{10} \frac{N(N-1)}{2} \rfloor \tag{3.63}$$

où $\lfloor \ \rfloor$ désigne l'arrondi inférieur.

Dans chaque classe, on calcule alors la *semi-variance*

$$\gamma(d_i) = \frac{1}{2N_i} \sum_{j=1}^{N} \sum_{k=1}^{N} \delta_{jk}^{i} (z_j - z_k)^2, \qquad i = 1, 2, \ldots, Nc \tag{3.64}$$

où d_i désigne le centre de la classe i, N_i le nombre de couples dans cette classe et où δ_{jk}^{i} vaut 1 si les points j et k appartiennent à la classe i et 0 dans le cas contraire.

La semi-variance γ constitue un mesure de l'erreur quadratique moyenne commise en estimant la valeur du champ à partir d'une observation effectuée à une distance d du point courant. En général, le variogramme est donc une fonction croissante de la distance d et présente une allure semblable à celle de la figure 3.5.

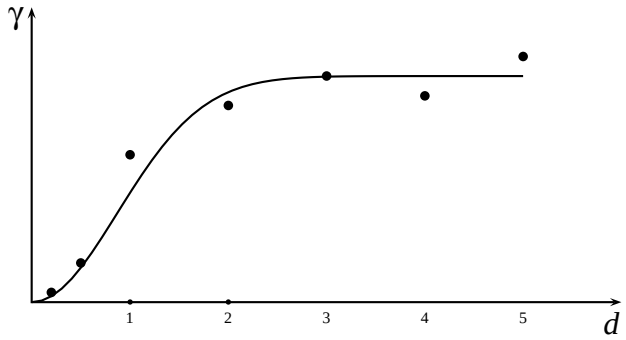


FIG. 3.5 – Semi-variogramme expérimental (points) et approximation par un modèle gaussien.

Idéalement, le variogramme possède une asymptote horizontale à un niveau qui correspond à la variance du champ analysé[3]. L'existence de cette asymptote montre que la variance du champ est finie et, surtout, que la fonction d'auto-correlation du champ ne dépend que de la distance entre les points et non pas de leur position dans le domaine étudié (stationnarité du second ordre). Cette propriété est évidemment capitale pour que le variogramme ait un sens et puisse être utilisé pour guider l'interpolation des données dans tous l'espace.

La valeur de la distance d à partir de laquelle l'asymptote horizontale est quasiment atteinte constitue une mesure de la distance maximale pour laquelle les données sont corrélées. C'est donc une estimation de la taille des structures les plus grandes présentent dans les données.

Normalement, $\gamma(d)$ tend vers zéro pour des distances d très petites. En pratique, on observe parfois une valeur non nulle. C'est ce que l'on appelle l'*effet pépite* ('nugget effect'). Celui-ci est dû aux erreurs de mesure ainsi qu'à l'échantillonnage qui ne permet pas de décrire les échelles spatiales les plus fines.

Le variogramme construit expérimentalement par une analyse par classe des données ne permet pas de décrire les variations continues de la semi-variance $\gamma(d)$. Il présente également des imperfections par rapport au modèle idéal de la figure 3.5. L'état suivante consiste donc en l'ajustement d'un modèle analytique au variogramme expérimental. Plusieurs types de modèles peuvent être utilisés :

i. modèle sphérique

$$\gamma(d) = \begin{cases} C_1 \left[1.5\dfrac{d}{a} - 0.5 \left(\dfrac{d}{a} \right)^3 \right] & \text{si} \quad d \leq a \\ C_1 & \text{si} \quad d > a \end{cases} \tag{3.65}$$

[3]L'analyse de données réelles par classes ne permet évidemment pas de décrire cette asymptote pour $d \to \infty$ ni même de l'approcher dans le cas où on dispose de peu de données expérimentales.

ii. modèle exponentiel

$$\gamma(d) = C_1 \left[1 - \mathrm{e}^{-3d/a}\right] \tag{3.66}$$

iii. modèle gaussien

$$\gamma(d) = C_1 \left[1 - \mathrm{e}^{-3d^2/a^2}\right] \tag{3.67}$$

iv. modèle à effet Hole

$$\gamma(d) = C_1 \left[1 - \frac{\sin d/a}{d/a}\right] \tag{3.68}$$

Ce modèle de Hole est appliqué aux variogrammes expérimentaux non monotones qui correspondent généralement à des champs présentant une structures spatiale périodique.

Ces modèles peuvent éventuellement être modifiés pour inclure un terme constant représentant l'effet pépite. Le choix d'un modèle plutôt qu'un autre est plus un art qu'une science... En toute rigueur, le choix d'un variogramme doit être validé par des tests statistiques.

Armé du modèle de variogramme, on procède à l'interpolation proprement dite. En chacun des points de la grille d'analyse, on déterminer les poids de telle façon que les semi-variances calculées à partir du point courant se retrouvent sur la courbe $\gamma(d)$ du modèle de variogramme choisi.

En chaque point de la grille d'analyse où le champ doit être déterminé, celui-ci est calculé selon

$$z_p = \sum_{i=1}^{N} w_i z_i \tag{3.69}$$

Cette relation étant linéaire, la cohérence du modèle de corrélation spatiale exige qu'une relation semblable existe entre les semi-variances, *i.e.*

$$\gamma(d_{pj}) = \sum_{i=1}^{N} w_i \gamma(d_{ij}), \qquad j = 1, 2, \ldots N \tag{3.70}$$

où d_{pj} désigne la distance entre le point d'analyse et le point de mesure j et d_{ij} désigne la distance entre les points de mesure i et j. En plus des relations (3.70), on souhaite que la valeur de l'interpolation constitue une combinaison convexe des données initiales, *i.e.*

$$w_1 + w_2 + w_3 + \cdots w_N = 1 \tag{3.71}$$

de façon à ne pas créer d'extrema en dehors du domaine des valeurs initiales. Les relations (3.70)-(3.71) constituent un système de $N+1$ équations pour les N poids inconnus w_i. On introduit alors une variable supplémentaire λ (multiplicateur de Lagrange) pour minimiser la variance de l'estimation. Les poids sont donc finalement déterminés en résolvant un système de la forme

$$\mathsf{C}\,\mathsf{w}_p = \mathsf{d}_p \tag{3.72}$$

où

$$\mathsf{C} = \begin{pmatrix} \gamma(d_{11}) & \gamma(d_{12}) & \cdots & \gamma(d_{1N}) & 1 \\ \gamma(d_{21}) & \gamma(d_{22}) & \cdots & \gamma(d_{2N}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(d_{N1}) & \gamma(d_{N2}) & \cdots & \gamma(d_{NN}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}, \qquad \mathsf{w}_p = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \\ \lambda \end{pmatrix}, \qquad \mathsf{d}_p = \begin{pmatrix} \gamma(d_{p1}) \\ \gamma(d_{p2}) \\ \vdots \\ \gamma(d_{pN}) \\ 1 \end{pmatrix} \tag{3.73}$$

en chaque point de la grille d'interpolation. Dans le krigeage, les poids sont donc de nature statistique plutôt que géométrique.

En plus du champ interpolé, le krigeage fournit une mesure de la variance des valeurs calculées en chaque point de la grille. Celle-ci est obtenue par la relation

$$\sigma_p^2 = \sigma_{data}^2 - \mathsf{w}_p{}^{\mathrm{T}} \mathsf{d}_p \tag{3.74}$$

où σ_{data}^2 désigne la variance des données initiales. Si on suppose que les erreurs d'estimation sont normalement distribuées, cette variance peut être calculée pour construire des intervalles de confiance autour des valeurs estimées. Par exemple, la probabilité que la valeur vraie se situe dans un intervalle $\pm\sigma_p$ (resp. $\pm 2\sigma_p$) autour de la valeur estimée est de 68 %(resp. 95%).

Pour estimer la robustesse de l'interpolation, on peut également avoir recours à une validation croisée consistant à écarter une observation de l'analyse et à comparer sa valeur avec l'estimation produite par application du krigeage aux données restantes. En répétant l'opération pour chacune des mesures considérées séparément, on peut calculer l'erreur quadratique moyenne ou le coefficient de correlation entre les données et leur estimation par krigeage.

Des raffinements de la méthode sont possibles. Par exemple, comme dans le cas de l'interpolation par distance inverse, on peut limiter le nombre de données intervenant dans la détermination des valeurs en un point donné.

On peut également traiter des champs anisotropes en construisant des variogrammes différents pour décrire les corrélations spatiales non seulement en fonction de la distance entre les points mais également en fonction de leurs positions relatives.

Il existe de nombreuses variantes de la méthode du krigeage. La méthode de base présentée ci-dessus est appelée *krigeage ordinaire*. Elle s'applique à des variables stationnaires de moyenne inconnue. La méthode du krigeage universel, par contre, peut être appliquée à des données non-stationnaires, *i.e.* qui contiennent une tendance. On peut également étendre le krigeage à plusieurs variables distinctes traitées simultanément (krigeage multivarié).

3.6 EOF.

L'interpolation optimale permet de ramener sur une grille régulière des données expérimentales distribuées de façon quelconque de façon. En même temps qu'elle fournit

une base rationnelle pour l'interpolation spatiale des mesures, la méthode introduit un certain lissage des données expérimentales en éliminant les structures qui paraissent peu ou pas fiables ou insuffisamment représentées par les mesures.

La décomposition en Fonctions Empiriques Orthogonales - EOF (*Empirical Orthogonal Functions*) poursuit un but semblable mais pour des données qui varient dans l'espace et dans le temps. Plus précisément, la décomposition en EOF vise à interpréter les données comme une superposition d'oscillations indépendantes. En général, un petit nombre de ces EOF suffisent à décrire l'essentiel de la variabilité spatiale et temporelle des données. Dès lors, ces quelques EOF les plus significatives fournissent une description compacte des données dans laquelle une grande partie du bruit expérimental (résultant d'erreurs de mesure ou de mesures peut significatives) a été éliminée. Dans certains cas, mais pas nécessairement, une explication dynamique peut être donnée aux EOF ainsi identifiées.

L'analyse EOF est connue sous le nom d'*analyse en composantes principales* en statistique pure et d'*analyse de facteurs* en sciences sociales. Dans tous les cas, il s'agit d'une méthode visant à réduire les données initiales pour faire apparaître les informations les plus significatives.

Considérons donc une série de données variable dans l'espace et dans le temps. Notons a_{ij} la mesure de la grandeur $a(\mathbf{x}_i, t_j)$ au point $\mathbf{x}_i$ $(i = 1, 2, \ldots, M)$ au temps t_j $(j = 1, 2, \ldots, N)$. Ces données peuvent être groupées dans une matrice A dont les colonnes décrivent l'état du système au temps t_j et les lignes représentent l'évolution temporelle au point $\mathbf{x}_i$. Le but de la procédure est de décrire les données sous la forme

$$a_{ij} = \sum_{k=1}^{M} \psi_i^{(k)} w_j^{(k)} \tag{3.75}$$

Dans cette expression, les $\psi_i^{(k)}$ $(k = 1, 2, \ldots, M)$ représentent M distributions spatiales particulières, *i.e.* M modes spatiaux ou EOF, et les $w_j^{(k)}$ introduisent une modulation temporelle des modes spatiaux. La modulation temporelle de chaque EOF est la même en tous les points du domaine. Au total, M modes sont utilisés pour représenter exactement les données initiales. De façon alternative, on peut interpréter (3.75) comme la modulation spatiale de M modes temporels d'oscillation.

La décomposition (3.75) des données pourrait a priori être réalisée en utilisant des modes spatiaux $\psi_i^{(k)}$ quelconques. Si les modes $\psi_i^{(k)}$ sont choisis de façon convenable, (3.75) permet de déterminer univoquement les $M \times N$ coefficients temporels $w_j^{(k)}$ correspondant aux $M \times N$ données a_{ij}. On désire cependant que ceux-ci soient liés aux données initiales et indépendants. Plus exactement, on souhaite que les modes spatiaux soient orthogonaux, *i.e.*

$$\sum_{i=1}^{M} \psi_i^{(k)} \psi_i^{(l)} = \delta_{kl} = \begin{cases} 1 & \text{si} \quad k = l \\ 0 & \text{si} \quad k \neq l \end{cases} \tag{3.76}$$

La relation (3.76) est la condition d'orthogonalité des vecteurs $\psi_\star^{(k)}$ et $\psi_\star^{(k)}$ correspondant aux modes k et l.

La condition (3.76) ne suffit évidemment pas à déterminer à elle seule les EOF. Pour aller plus loin, une condition similaire est demandée aux coefficients temporels, soit

$$\frac{1}{N}\sum_{j=1}^{N} w_j^{(k)} w_j^{(l)} = \sigma_k^2 \delta_{kl} = \begin{cases} \sigma_k^2 & \text{si} \quad k = l \\ 0 & \text{si} \quad k \neq l \end{cases} \tag{3.77}$$

Cette condition exprime que les facteurs temporels relatifs à des modes différents sont non corrélés entre-eux. Pour $k = l$, σ_k^2 représente la variance associée au mode k.

Les conditions (3.76) et (3.77) suffisent pour déterminer les EOF et les coefficients temporels. Pour le voir, ré-écrivons d'abord (3.75) sous la forme matricielle équivalente

$$\mathsf{A} = \mathsf{X}\mathsf{W}^{\mathrm{T}} \tag{3.78}$$

où

$$\mathsf{X} = \begin{pmatrix} \psi_\star^{(1)} & \psi_\star^{(2)} & \cdots & \psi_\star^{(M)} \end{pmatrix} \tag{3.79}$$

et

$$\mathsf{W} = \begin{pmatrix} w_\star^{(1)} & w_\star^{(2)} & \cdots & w_\star^{(M)} \end{pmatrix} \tag{3.80}$$

Avec ces notations, les conditions (3.76) et (3.77) s'écrivent respectivement

$$\mathsf{X}^{\mathrm{T}}\mathsf{X} = \mathbb{I}, \qquad \mathsf{W}^{\mathrm{T}}\mathsf{W} = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\right) \tag{3.81}$$

On calcule dès lors aisément

$$\mathsf{A}\mathsf{A}^{\mathrm{T}}\mathsf{X} = \mathsf{X}\mathsf{W}^{\mathrm{T}}\mathsf{W}\mathsf{X}^{\mathrm{T}}\mathsf{X} = \mathsf{X}\mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\right) \tag{3.82}$$

ce qui montre que les colonnes de X, *i.e.* les EOF $\psi_\star^{(k)}$ sont les vecteurs propres de la matrice $\mathsf{A}\mathsf{A}^{\mathrm{T}}$. Les valeurs propres sont les variances σ_k^2 des modes correspondants.

De même,

$$\mathsf{A}^{\mathrm{T}}\mathsf{A}\mathsf{W} = \mathsf{W}\mathsf{X}^{\mathrm{T}}\mathsf{X}\mathsf{W}^{\mathrm{T}}\mathsf{W} = \mathsf{W}\mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_M^2\right) \tag{3.83}$$

Les facteurs temporels sont donc les vecteurs propres de la matrice $\mathsf{A}^{\mathrm{T}}\mathsf{A}$. Les deux matrices $\mathsf{A}^{\mathrm{T}}\mathsf{A}$ et $\mathsf{A}\mathsf{A}^{\mathrm{T}}$ sont symétrique et semi-définies positives. Elles partagent les mêmes valeurs propres non nulles σ_k^2.

En général, on numérote les EOF par valeur décroissante de la variance σ_k^2. La première EOF représente donc le signal le plus énergétique. La seconde EOF, représente le mode orthogonal au premier et décrivant la plus grande partie de la variance restante... La somme des valeurs propres σ_k^2 est égale à la variance totale des données initiales.

En pratique, l'identification des EOF est généralement réalisée en s'appuyant sur la *décomposition en valeurs singulières* de la matrice A. On montre, en effet, que toute matrice réelle A de dimensions $M \times N$ peut être décomposée en un produit

$$\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{V}^{\mathrm{T}} \tag{3.84}$$

où U et V sont des matrices orthogonales ($\mathsf{U}^{\mathrm{T}}\mathsf{U} = \mathsf{I}$, $\mathsf{V}^{\mathrm{T}}\mathsf{U} = \mathsf{I}$) respectivement d'ordre M et N et où D est une matrice de dimensions $M \times N$ dont tous les éléments sont nuls à l'exception des éléments $\mathsf{D}_{kk} = \sigma_k$ où $\sigma_k > 0$ sont appelés les *valeurs singulières* de A. Les colonnes des matrices U et V sont, respectivement, les vecteurs propres orthonormés de $\mathsf{A}\mathsf{A}^{\mathrm{T}}$ et $\mathsf{A}^{\mathrm{T}}\mathsf{A}$ relatifs aux valeurs propres non nulles σ_k^2 qui sont communes aux deux matrices. En comparant (3.84) à (3.78), on constate que les matrices X et W recherchées coïncident avec les matrices U et $\mathsf{V}\mathsf{D}^{\mathrm{T}}$ fournies par la décomposition en valeurs singulières de la matrice A.

Pour illustrer la capacité de l'analyse EOF a filtrer les données bruitées, considérons le champ du type

$$C(t,x,y) = 7A(t)\cos\frac{2\pi x}{L_x}\cos\frac{2\pi y}{L_y} + 1B(t)\cos\frac{4\pi x}{L_x}\cos\frac{4\pi y}{L_y} + 0.5\varepsilon(t,x,y) \tag{3.85}$$

où A, B et ε désignent des variables aléatoires issues d'une distribution normale de moyenne nulle et de variance unitaire. En chaque temps intermédiaire, le champ total est la composition de deux distributions de base et d'un bruit aléatoire. La figure 3.6 montre une vue particulière de ce champ en un instant particulier.

L'application de l'analyse EOF permet de retrouver les distributions de base

$$\cos\frac{2\pi x}{L_x}\cos\frac{2\pi y}{L_y} \tag{3.86}$$

et

$$\cos\frac{4\pi x}{L_x}\cos\frac{4\pi y}{L_y} \tag{3.87}$$

comme première et deuxième EOF (Figure 3.7). La troisième EOF ne présente aucune structure particulière : elle correspond au bruit inclus dans (3.85). Remarquons que, puisque les EOF sont normalisées selon(3.76), les valeurs absolues des EOF n'ont aucune signification physique.

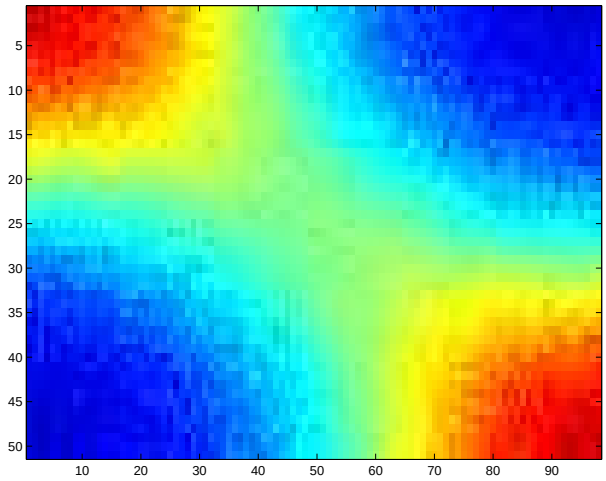


FIG. 3.6 – Vue du champ (3.85) en un instant particulier.

Les calcul des carrés des valeurs singulières (éléments diagonaux non nuls de D) permet de déterminer la variance expliquée par chacun des modes. Dans le cas étudié, les deux premiers modes suffisent à décrire le champ initial (Fig. 3.8).

On peut donc éliminer le bruit du champ initial en recombinant les deux premières EOF. Le résultat de cette opération, présenté à la figure 3.9, peut être comparé à la figure initiale (Fig. 3.6).

Dans l'exemple traité, les fonctions (3.86) et (3.87) utilisées pour construire le champ (3.85) sont orthogonales. L'analyse par EOF permet donc de les retrouver telles qu'elles. Dans le cas général, l'analyse EOF tente de construire le champ comme la superposition de modes orthogonaux, même si celui-ci pourrait être représenté de façon plus compacte par des modes non orthogonaux. Dans ce cas, les modes non orthogonaux sont répartis dans les EOF successives et peuvent ne pas s'identifier à une EOF précise. Si ceci à pour effet de répartir la variance totale en un plus grand nombre de modes, le champ reste cependant généralement décrit par un nombre limité d'EOF : les valeurs singulières décroissant rapidement avec le numéro des EOF.

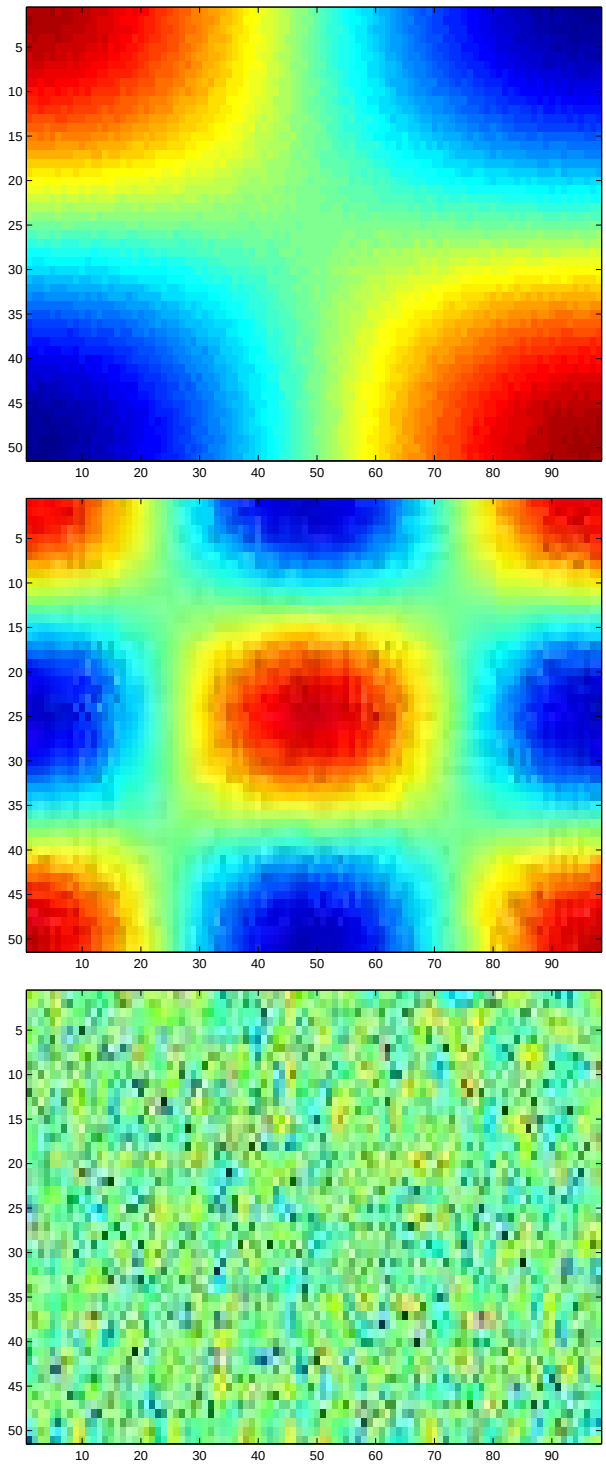


FIG. 3.7 – EOF n° 1, 2 et 3 du champ (3.85)

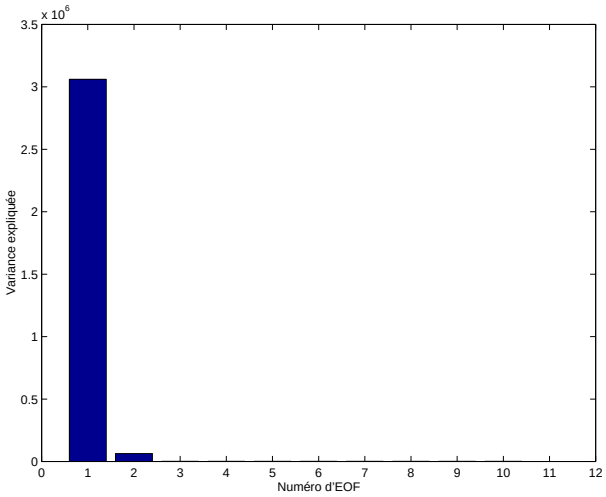


FIG. 3.8 – Variance expliquée par les premières EOF.

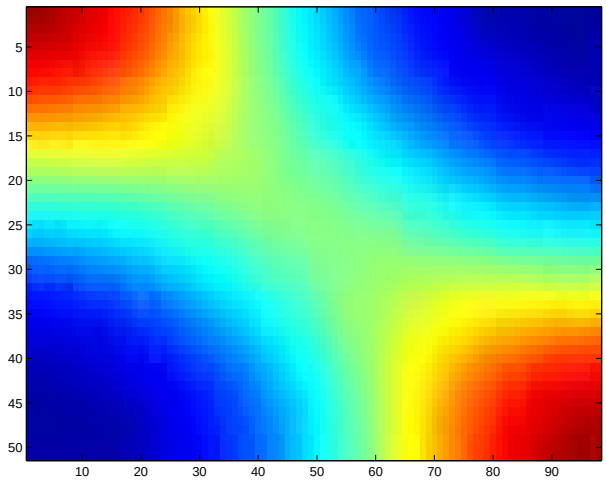


FIG. 3.9 – Reconstruction du champ de la figure 3.6 à partir des deux premières EOF.

Chapitre 4

Analyse de séries temporelles

Les méthodes modernes d'observation et d'enregistrement de données fournissent de longues séries temporelles. L'un des buts de l'analyse de ces données est généralement de mettre en évidence la variabilité ou la structure sous-jacente pour mieux comprendre les phénomènes impliqués. L'expérimentateur pourra ainsi s'attacher à identifier les fréquences dominantes et a séparer le signal principal des fluctuations associées au bruit ou aux erreurs de mesure.

Les techniques utilisées reposent généralement sur l'hypothèse d'*ergodicité, i.e.* on suppose que les propriétés statistiques des séries temporelles sont indépendantes du temps. Dans ce cas, les moyennes temporelles des données sont équivalentes à des moyennes d'ensemble et peuvent être rattachées aux méthodes statistiques standard.

4.1 Concepts de base.

Considérons une série de N valeurs $\{y_1, y_2, \ldots, y_n\}$ mesurées aux temps (discrets) $\{t_1, t_2, \ldots, t_n\}$. On supposera ici que les instants successifs sont séparés d'un pas Δt constant. Pour caractériser cette série, on peut commencer par calculer

– la *moyenne* μ, donnée par

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{4.1}$$

– la *variance* σ^2, donnée par[1]

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} \left[y_i - \mu\right]^2 \tag{4.2}$$

[1] On notera que la variance est ici calculée en divisant par N la somme des carrés des écarts à la moyenne. Cette expression est applicable lorsque la moyenne est connue indépendamment des données de la série dont on calcule la variance. En général, lorsque la moyenne et la variance sont estimées à partir des mêmes données, un facteur $1/(N-1)$ plutôt que $1/N$ est introduit dans le calcul de la variance pour obtenir un meilleur estimateur. Dans le contexte de la présente analyse, il importe cependant de s'en tenir à la définition (4.2).

La racine carrée de la variance est l'*écart-type*.

Pour caractériser le degré de stabilité temporelle du signal décrit par la série de données, on peut comparer les valeurs prises en des instants successifs $t_1 = t$ et $t_2 = t + \tau$ pour différentes valeurs du décalage temporel τ. La comparaison correspondante des données brutes conduit à la définition de la fonction d'*auto-correlation*

$$R_{yy}(\tau) = \frac{1}{N-k} \sum_{i=1}^{N-k} y_i y_{i+k} \tag{4.3}$$

où $\tau = k\Delta t$ $(k = 0, 1, \ldots, M)$. Comme on le voit, ce calcul fait intervenir un nombre d'autant plus restreint de termes que k est grand. Pour que la somme (4.3) conserve son sens statistique, il convient de prendre $M \ll N$.

Un calcul semblable peut être mené à partir de la série de données dont on a préalablement soustrait la moyenne. On définit ainsi la fonction d'*auto-covariance*

$$C_{yy}(\tau) = \frac{1}{N-k} \sum_{i=1}^{N-k} (y_i - \mu)(y_{i+k} - \mu) \tag{4.4}$$

Pour un décalage $\tau = 0$, on calcule aisément

$$C_{yy}(0) = \sigma^2 = R_{yy}(0) - \mu^2 \tag{4.5}$$

Une division de l'auto-covariance par la variance conduit alors naturellement à la définition de la fonction d'*auto-covariance normalisée*

$$\rho_{yy}(\tau) = \frac{C_{yy}(\tau)}{\sigma^2} \tag{4.6}$$

Par construction, on a $|\rho_{yy}(\tau)| \leq 1$ pour tout τ et $\rho_{yy}(0) = 1$.

Les fonctions d'auto-corrélation et auto-covariance permettent de quantifier le degré de stabilité, ou au contraire de variabilité, d'un signal. En effet, une valeur élevée, *i.e.* proche de l'unité, de la fonction d'auto-covariance normalisée pour un délai τ donné, indique que le signal ne change pas fondamentalement entre deux instants séparés de τ ou, plus généralement, que l'observation des valeurs prises en un instant t permet une bonne prévision des valeurs en $t + \tau$. Pour caractériser l'échelle de temps de cette auto-corrélation, on pourra dés lors utiliser le *temps caractéristique intégral*

$$T^\star = \frac{\Delta\tau}{2} \sum_{i=0}^{N'} [\rho_{yy}(\tau_i) + \rho_{yy}(\tau_{i+1})] = \frac{\Delta\tau}{2\sigma^2} \sum_{i=0}^{N'} [C_{yy}(\tau_i) + C_{yy}(\tau_{i+1})] \tag{4.7}$$

où $N' \leq N - 1$ est tel que la somme dans (6.39) converge vers une valeur constante. Si la somme ne se stabilise pas, on en conclut que la série ne possède pas de temps caractéristique intégral ou on limite la somme jusqu'à la première annulation de la fonction d'auto-corrélation.

Lorsqu'il existe, le temps caractéristique $T^\star$ est tel que les données peuvent être considérées comme non corrélées pour des écarts supérieurs à $T^\star$. Dès lors, le nombre de degrés de liberté de la série est approximativement donné par $N\Delta t/T^\star$.

À titre d'exemple, considérons d'abord une série temporelle $\varepsilon(t)$ purement aléatoire (bruit blanc) dont les valeurs sont choisies selon une loi de probabilité normale (de moyenne μ_0 et de variance σ_0^2) et dont les valeurs successives sont indépendantes. La moyenne et la variance de la série temporelle sont alors égales à celles de la loi de probabilité dont sont issues les valeurs de la série. De plus, on a

$$R_{\varepsilon\varepsilon}(\tau) = \sigma_0^2 \rho_{\varepsilon\varepsilon}(\tau) = \begin{cases} \sigma_0^2 & \text{pour } \tau = 0 \\ 0 & \text{sinon.} \end{cases} \tag{4.8}$$

L'annulation de la fonction d'auto-corrélation pour tout délai non nul montre bien l'indépendance des valeurs successives.

Considérons maintenant le signal purement périodique décrit par

$$y_i = A \sin \frac{2\pi i \Delta t}{T}, \qquad (i = 1, 2, \ldots)$$

où la période $T/\Delta t$ est entier. La moyenne du signal étant nulle, les fonctions d'auto-corrélation et d'auto-covariance sont égales. On calcule successivement

$$\sigma^2 = \frac{1}{2}$$

et

$$\rho(\tau) = \cos \frac{2\pi\tau}{T}$$

de sorte que la fonction d'auto-corrélation normalisée présente la même périodicité que la série originelle. Elle indique bien que les enregistrement séparés d'un nombre entier de périodes sont parfaitement corrélés alors qu'un valeur ($\rho = -1$) caractérise l'opposition de phase pour un décalage temporelle d'une demi-période.

4.2 Séries de Fourier.

L'un des buts de l'analyse des séries temporelles est de mettre en évidence les périodes caractéristiques contenues dans le signal étudié. La résolution de ces questions trouve son origine dans la théorie des séries et des transformations de Fourier que nous allons donc effleurer ici.

Les concepts fondamentaux et toutes leurs extensions sont contenues dans le résultat de base qui dit que toute fonction périodique suffisamment régulière[2] peut être représenté comme une série, *i.e.* une somme infinie, de fonctions sinus et cosinus dont les périodes

[2] Il suffit que la fonction et sa dérivées soient continues par morceaux. Ces conditions sont connues sous le nom de conditions de Dirichlet.

sont les sous-multiples de la période T du signal périodique global. Mathématiquement, si le signal $f(t)$ présente une période T, on aura

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos\left(\frac{2\pi k t}{T}\right) + b_k \sin\left(\frac{2\pi k t}{T}\right) \right] \tag{4.9}$$

où les coefficients a_0, a_k et b_k sont les coefficients de Fourier de f. La décomposition du signal en ses composantes harmoniques élémentaires (4.9) trouve son écho également dans la relation de Parseval

$$\frac{1}{T}\int_{t_0}^{t_0+T} |f(t)|^2 dt = \frac{a_0^2}{4} + \frac{1}{2}\sum_{k=1}^{\infty}(a_k^2 + b_k^2) \tag{4.10}$$

Quand on sait que le carré d'un signal est généralement considéré comme représentatif de l'énergie contenue dans celui-ci, la relation de Parseval (4.10) signifie donc que l'énergie totale (la moyenne de cette énergie sur une période) peut être calculée simplement comme la somme des énergies des composantes harmoniques élémentaires considérées séparément.

La relation de Parseval résulte de l'indépendance relative des composantes harmoniques en lesquelles le signal de départ a été décomposé. Mathématiquement, cette indépendance témoigne de l'orthogonalité entre les fonctions sinus et cosinus considérées, *i.e.*

$$\frac{1}{T}\int_{t_0}^{t_0+T} \sin\left(\frac{2\pi k t}{T}\right)\cos\left(\frac{2\pi \ell t}{T}\right) dt = 0 \qquad \forall k, \ell \tag{4.11}$$

$$\frac{1}{T}\int_{t_0}^{t_0+T} \cos\left(\frac{2\pi k t}{T}\right)\cos\left(\frac{2\pi \ell t}{T}\right) dt = 0 = \begin{cases} 1 & \text{si } k = \ell = 0 \\ \frac{1}{2} & \text{si } k = \ell > 0 \\ 0 & \text{si } k \neq \ell \end{cases} \tag{4.12}$$

$$\frac{1}{T}\int_{t_0}^{t_0+T} \sin\left(\frac{2\pi k t}{T}\right)\sin\left(\frac{2\pi \ell t}{T}\right) dt = 0 = \begin{cases} 0 & \text{si } k = \ell = 0 \\ \frac{1}{2} & \text{si } k = \ell > 0 \\ 0 & \text{si } k \neq \ell \end{cases} \tag{4.13}$$

Les coefficients de Fourier d'une fonction suffisamment régulière peuvent être calculés en se basant sur ces relations d'orthogonalité. Par exemple, en multipliant les deux membre de (4.9) par $\sin(2\pi\ell/T)$ et en intégrant sur une période, il vient, en utilisant (4.11)-(4.13),

$$\begin{aligned} \int_{t_0}^{t_0+T} f(t)\sin\left(\frac{2\pi \ell t}{T}\right) dt &= \frac{a_0}{2}\int_{t_0}^{t_0+T} \sin\left(\frac{2\pi \ell t}{T}\right) dt \\ &\quad + \sum_{k=1}^{\infty} a_k \int_{t_0}^{t_0+T} \cos\left(\frac{2\pi k t}{T}\right)\sin\left(\frac{2\pi \ell t}{T}\right) dt \\ &\quad + \sum_{k=1}^{\infty} b_k \int_{t_0}^{t_0+T} \sin\left(\frac{2\pi k t}{T}\right)\sin\left(\frac{2\pi \ell t}{T}\right) dt \\ &= b_\ell \frac{T}{2} \end{aligned} \tag{4.14}$$

En utilisant cette procédure, on obtient donc les coefficients

$$\begin{cases} a_k = \dfrac{2}{T}\displaystyle\int_{t_0}^{t_0+T} f(t)\cos\left(\frac{2\pi kt}{T}\right)dt \\ b_k = \dfrac{2}{T}\displaystyle\int_{t_0}^{t_0+T} f(t)\sin\left(\frac{2\pi kt}{T}\right)dt \end{cases} \tag{4.15}$$

EXEMPLE 4.1 Développons en série de Fourier la fonction périodique f telle que $f(t) = 2|t|/T$ pour $t \in [-T/2, T/2]$ (qui prend donc la valeur 1 pour $t = \pm T/2$) et qui se prolonge en une fonction de période T.

En appliquant les formules (4.15), on montre aisément que les coefficients b_k sont tous nuls et que les coefficients a_k sont donnés par

$$\begin{aligned} a_0 &= \frac{2}{T}\int_{-T/2}^{T/2} f(t)dt = \frac{8}{T^2}\int_0^{T/2} t\,dt = 1 \\ a_k &= \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos\left(\frac{2\pi kt}{T}\right)dt = \frac{8}{T^2}\int_0^{T/2} t\cos\left(\frac{2\pi kt}{T}\right)dt \\ &= -\frac{2}{k^2\pi^2}(1-(-1)^k) = \begin{cases} 0 & \text{si } k \text{ est pair,} \\ -\dfrac{4}{\pi^2 k^2} & \text{si } k \text{ est impair.} \end{cases} \end{aligned}$$

Dès lors,

$$f(t) = \frac{1}{2} - \sum_{k=0}^{\infty} \frac{4}{\pi^2(2k+1)^2}\cos(2k+1)\omega t$$

où $\omega = 2\pi/T$ est la pulsation fondamentale. On constate sur la figure 4.1 que le signal de départ est de mieux en mieux approché à mesure que l'on augmente le nombre de composantes harmoniques.

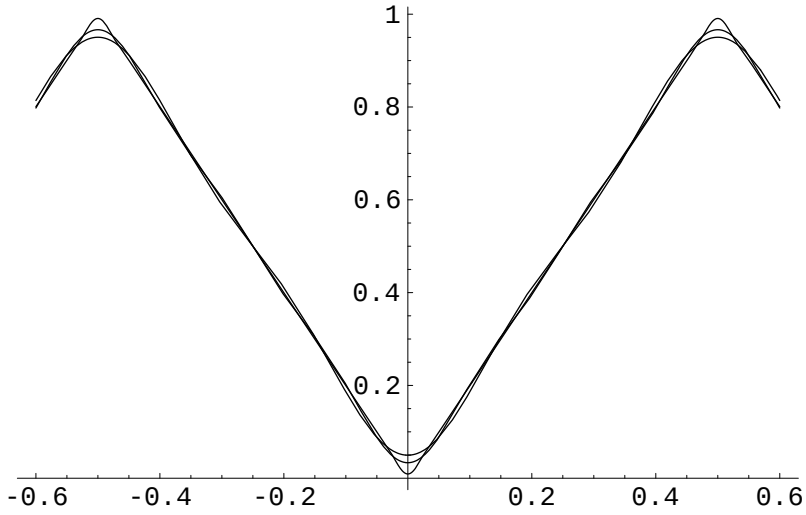


FIG. 4.1 – Reconstruction d'un signal périodique par composition de composantes élémentaires avec 1, 2 et 10 composantes de Fourier.

Remarquons que la série obtenue ne contient que des fonctions cosinus. C'est une conséquence de la parité de la fonction f. Inversement, la série de Fourier représentant une fonction impaire ne comporte que des composantes en sinus, des signaux élémentaires impairs.

◇

4.3 Transformée de Fourier.

La représentation d'un signal périodique par le biais d'un série de Fourier met en évidence la possibilité de caractériser le signal par les amplitudes des harmoniques qui interviennent dans sa construction. Dans le cas d'un signal périodique, les pulsations (par abus de language, on parlera souvent des fréquences) qui interviennent sont

$$0, \quad \frac{2\pi}{T}, \quad \frac{4\pi}{T}, \quad \frac{6\pi}{T}, \quad \frac{8\pi}{T}, \quad \cdots$$

Il s'agit donc d'un ensemble discret, mais infini, de pulsations séparées de $\Delta\omega = 2\pi/T$.

Interprétant une fonction quelconque, non périodique, comme une fonction périodique dont la période tendrait vers l'infini, la représentation en série de Fourier correspondante est caractérisé par une infinite de fréquences dont le 'saut' $\Delta\omega$ tend vers zéro. En d'autres termes, la somme (infinie) représentant la série de Fourier se transforme en une intégrale sur un spectre continu de fréquences. Ceci conduit donc à la définition de la transformée

de Fourier $\tilde{f}(\omega)$ d'une fonction $f(t)$ quelconque par[3]

$$\tilde{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, \mathrm{e}^{-i\omega t}\, dt \tag{4.16}$$

et la transformée inverse par

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(\omega)\, \mathrm{e}^{i\omega t}\, d\omega \tag{4.17}$$

Si on se souvient que l'exponentielle imaginaire est donnée par

$$\mathrm{e}^{i\omega t} = \cos\omega t + i \sin\omega t \tag{4.18}$$

on remarque que la transformée de Fourier $\tilde{f}(\omega)$ joue un rôle analogue aux coefficients de Fourier dans la représentation (4.17) du signal puisqu'elle pondère l'influence de chaque composante élémentaire de pulsation ω dans le signal total. Le calcul de la transformée de Fourier est également semblable à celui des coefficients de Fourier (4.15) puisque la transformée est obtenue en intégrant le produit du signal de départ et les fonctions de bases $\sin\omega t$ et $\cos\omega t$ combinées en une exponentielle imaginaire

$$\mathrm{e}^{-i\omega t} = \cos\omega t - i \sin\omega t \tag{4.19}$$

On sera attentif au fait que la transformée de Fourier définie par (4.16) est, en général, complexe. Si, comme c'est généralement le cas, la fonction $f(t)$ est réelle, alors

$$\tilde{f}(-\omega) = \overline{\tilde{f}(\omega)} \tag{4.20}$$

où $\bar{}$ désigne le complexe conjugué d'un nombre complexe. En combinant les contributions des pulsations $-\omega$ et ω dans (4.17), on trouve

$$\begin{aligned} \overline{\tilde{f}(\omega)}\, \mathrm{e}^{-i\omega t} + \tilde{f}(\omega)\, \mathrm{e}^{i\omega t} &= \overline{\tilde{f}(\omega)\, \mathrm{e}^{i\omega t}} + \tilde{f}(\omega)\, \mathrm{e}^{i\omega t} \\ &= 2\Re\left[\tilde{f}(\omega)\, \mathrm{e}^{i\omega t}\right] = 2\Re[\tilde{f}(\omega)] \cos\omega t - 2\Im[\tilde{f}(\omega)] \sin\omega t \\ &= 2|\tilde{f}(\omega)| \cos(\omega t + \phi) \end{aligned} \tag{4.21}$$

où $\Re$, $\Im$ et $|\cdot|$ désignent respectivement la partie réelle, la partie imaginaire et le module d'un nombre complexe et où $\phi = \arg(\tilde{f}(\omega))$ est l'argument de $\tilde{f}(\omega)$. Injectant cette expression dans (4.17), le signal de départ peut donc s'écrire sous la forme

$$f(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} |\tilde{f}(\omega)| \cos(\omega t + \arg \tilde{f}(\omega)) d\omega \tag{4.22}$$

[3]Le facteur $1/\sqrt{2\pi}$ introduit dans les définitions de la transformée de Fourier et de son inverse relève d'un choix particulier qui permet de maintenir une symétrie entre les deux expressions. Certains auteurs introduisent des facteurs différents dans les définitions de la transformée et de son inverse. Le produit de ces facteurs doit cependant être égal à $1/(2\pi)$.

On en déduit que le module de la transformée de Fourier $\tilde{f}(\omega)$ mesure l'amplitude des signaux harmoniques élémentaires alors que son argument décrit leurs déphasages respectifs.

La relation de Parseval (4.10) peut également être étendue au cas des fonctions non périodiques qui admettent une transformée de Fourier. On a, cette fois

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} |\tilde{f}(\omega)|^2 d\omega \tag{4.23}$$

qui indique que l'énergie du système peut indifféremment être évaluée à partir du signal dans le domaine temporel ou à partir de sa transformée de Fourier, dans le domaine fréquentiel. La quantité $|\tilde{f}(\omega)|^2$ décrit donc le spectre d'énergie du système étudié et permet d'identifier les fréquences dominantes dans le signal étudiée. Dans le cas particulier d'un signal périodique purement harmonique de pulsation ω, la transformée de Fourier est identiquement nulle sauf en $\pm\omega$ où elle présente deux pulses (impulsions de Dirac).

4.4 Transformée de Fourier discrète.

Dans le cadre expérimental habituel, les données à traiter ne sont pas représentées sous la forme de fonctions continues mais sous celle d'un ensemble fini de données discrètes que nous supposerons espacées d'un intervalle Δt constant. Les concepts introduits dans les sections précédentes doivent donc être généralisée à cette situation pratique.

Soit $x_0, x_1, \ldots, x_{N-1}$ les données à analyser. On appelle *transformée de Fourier Discrète* (en anglais DFT) la suite des nombres (complexes) $X_0, X_1, \ldots, X_{N-1}$ donnée par

$$X_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \, \mathrm{e}^{-\frac{2\pi i}{N} kj}, \qquad k = 0, 1, \ldots, N-1 \tag{4.24}$$

La *t*ransformée de Fourier discrète inverse (en anglais IDFT) est quant-à-elle définie[4] par

$$x_j = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k \, \mathrm{e}^{\frac{2\pi i}{N} kj}, \qquad j = 0, 1, \ldots, N-1 \tag{4.25}$$

On remarquera le parallélisme parfait entre ces formules et les expressions (4.16) et (4.17) correspondant au cas continu. En particulier, (4.25) montre que la série de données expérimentales peut s'écrire comme la somme d'une composante constante (pour $k = 0$) et d'exponentielles imaginaires, c'est-à-dire de signaux harmoniques, dont les pulsations sont des multiples de la pulsation $2\pi/N$ correspondant à la longueur totale du signal. Les

[4]Ici encore les facteurs $1/\sqrt{N}$ introduits relèvent d'un choix particulier induisant des expressions symétriques des deux transformées. Des facteurs différents peuvent être introduits dans les expression de la transformée discrète et de son inverse. Il importe seulement que le produit de ces facteurs soit égal à $1/N$.

nombres complexes X_k représentent l'amplitude et la phase des différentes composantes harmoniques.

La suite des données et sa transformée de Fourier discrète contiennent la même information ; elles sont toutes caractérisées par la donnée de N nombres correspondant, aux valeurs instantanées en des instants différents dans le cas de série initiale (domaine temporel), ou aux amplitude des composantes harmoniques dans le cas de la transformée de Fourier discrète (dans le domaine fréquentiel). Dans le cas le plus fréquent où les données initiales sont réelles, on a

$$X_k = \overline{X_{N-k}} \tag{4.26}$$

où la barre indique le nombre complexe conjugué. Dans ce cas, la moitié des coefficients de la transformée de Fourier suffit à décrire complètement celle-ci. Plus précisément, si le nombre de données N est pair[5], les coefficients X_0 et $X_{N/2}$ sont réels et les coefficients X_1, X_2, …, $X_{N/2}$ permettent à eux-seuls de représenter les données sous la forme d'une somme de fonctions sinus et cosinus. Le coefficient X_0 représente alors la composante constante du signal (au facteur $\sqrt{N}$ près) tandis que les autres coefficients de Fourier caractérisent l'importance des différentes composantes harmoniques de pulsations

$$\omega_1 = \frac{2\pi}{N\Delta t}, \qquad \omega_k = k\omega_1 \qquad k = 1, 2, \ldots, N/2 \tag{4.27}$$

Les données initiales x_j peuvent être considérées comme provenant de l'échantillonnage aux instants $t_j = j\Delta t$ $(j = 0, 1, \ldots, N-1)$ du signal continu $x(t)$ tel que

$$x(t) = \frac{1}{\sqrt{N}} \left[X_0 + 2 \sum_{k=1}^{N/2-1} |X_k| \cos(k\omega_1 t + \phi_k) + X_{N/2} \cos \frac{N\omega_1 t}{2} \right] \tag{4.28}$$

où $\phi_k = \arg X_k$.

L'écriture (4.28) fait clairement apparaître l'influence des paramètres critiques que sont la longueur totale de la série de données et l'intervalle d'échantillonnage.

Le dernier terme de (4.28) correspond à la plus haute fréquence pouvant être décrite par un échantillonnage à un intervalle de temps Δt. La fréquence correspondante est appelée la *fréquence de Nyquist*

$$f_c = \frac{1}{2\Delta t}, \qquad \omega_c = 2\pi f_c = \frac{\pi}{\Delta t}. \tag{4.29}$$

Elle correspond à des oscillations de période $2\Delta t$. L'intervalle de temps entre deux données successives détermine dont la plus grande fréquence accessible à l'analyse.

La pulsation fondamentale ω_1 correspond à des signaux dont la période est égale à la durée totale de la série de mesures. Cette pulsation fondamentale correspond également

[5] En pratique, la transformée de Fourier discrète peut être évaluée très efficacement par l'algorithme de la *Transformée de Fourier Rapide (FFT)*. Cette méthode n'est applicable que si le nombre de données est une puissance de 2. Le nombre de données analysées sera donc bien généralement pair.

à la résolution $\Delta\omega = 2\pi/(N\Delta t)$ dans le domaine fréquentiel. Dés lors, si on augmente la longueur de l'enregistrement, sans changer le Δt, on obtiendra une résolution plus grande dans le domaine fréquentiel et on pourra distinguer entre-elles des composantes dont les pulsations sont plus proches l'une de l'autre.

La relation de Parseval se généralise également sans problème. Elle prend ici la forme

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |x_j|^2 = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} |X_k|^2 \tag{4.30}$$

Dans le cas habituel de données réelles décrites par un nombre pair de données, on peut également écrire

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |x_j|^2 = \frac{1}{\sqrt{N}} \left[|X_0|^2 + 2 \sum_{k=1}^{N/2-1} |X_k|^2 + |X_{N/2}|^2 \right] \tag{4.31}$$

où les différents termes de la somme sont représentatifs de l'énergie contenue dans les différents modes. En réalité, cette énergie est caractéristique des processus dont la pulsation est comprise dans une bande de largeur $\Delta\omega$ autour de la pulsation nominale ω_k. Dès lors, il est d'usage de normaliser ces coefficients et de définir la densité spectrale en divisant par $\Delta\omega$,

$$S(\omega_k) = \frac{2|X_k|^2}{\Delta\omega}, \qquad k = 1, 2, \ldots \tag{4.32}$$

EXEMPLE 4.2 À titre d'exemple, considérons les données représentées à la figure 4.2 obtenues en échantillonnant un signal inconnu en $N = 32$ instants successifs séparés de temps $\Delta t = 0.375s$. La durée totale de l'enregistrement est donc de $T = 12\ s$.

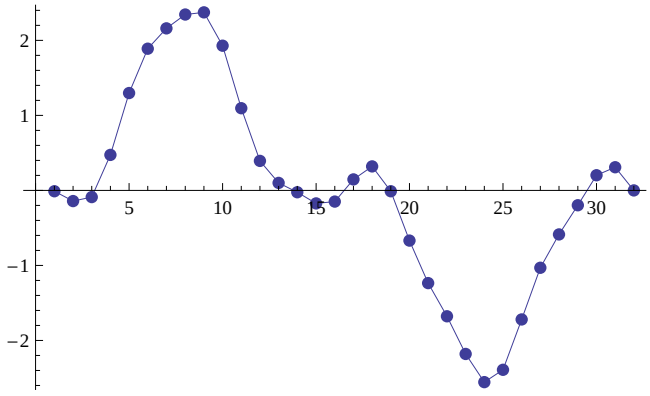


FIG. 4.2 – Données mesurées

L'application de la procédure décrite ci-dessus permet de calculer la transformée de Fourier discrète dont les modules des $N/2$ premiers termes sont représentés à la figure 4.2.

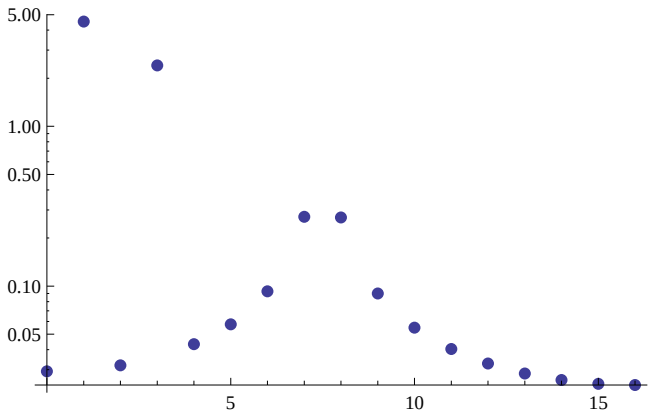


FIG. 4.3 – Module des coefficients de Fourier de la série représentée à la figure 4.2 (échelle logarithmique).

La figure 4.2 fait clairement apparaître que la transformée de Fourier discrète ne comporte que 4 termes significatifs correspondant à X_1, X_3, X_7 et X_8. On en déduit que le signal est dominé par les composantes harmoniques dont les périodes sont données par

$$T_1 = T = 12\ s, \qquad T_3 = 4\ s, \qquad T_7 = 12/7\ s, \qquad T_8 = 12/8\ s$$

La présence de termes significatifs de fréquences proches comme T_7 et T_8 suggère l'existence d'une composante de fréquence intermédiaire à celles de T_7 et T_8 qui ne peut être correctement décrite par l'échantillonnage disponible. En réalité, les données représentées à la figure 4.2 ont été obtenues en échantillonnant le signal

$$y(t) = 1.6 \sin \frac{2\pi t}{12} - 0.85 \sin \frac{2\pi t}{4} + 0.15 \sin \frac{2\pi t}{1.6}$$

À l'évidence, les composantes de période 12 et 4 secondes sont parfaitement captées par l'analyse, à l'inverse de la composante de période 1.6 secondes. Cette dernière est donc artificiellement répartie sur les périodes de bases comprises dans l'analyse discrète de Fourier. Un signal de période $T = 1.5\ s$ aurait par contre été décrit parfaitement puisqu'il correspondrait à la composante $k = 12/1.5 + 1 = 9$.

◇

4.5 Filtrage.

4.5.1 Principe général.

Très souvent les séries temporelles obtenues expérimentalement présentent des variations rapides et apparemment erratiques qui masquent ou obscurcissent le signal principal que l'on souhaite étudier. Ainsi, des variations journalières se superposent

au signal saisonnier dans l'enregistrement de l'insolation en un point. De même, les mesures courantométriques ne fourniront pas une image claire des oscillations de marée mais seront perturbées par des variations à plus haute fréquence induites par les coups de vents successifs, les vagues et la houle. Dans ce contexte, le but du filtrage est de 'purifier' l'enregistrement expérimental en mettant en évidence le *signal* utile et en éliminant les variations aux hautes fréquences qui sont dès lors considérées comme un *bruit* de fond indésirable. Remarquons que l'interprétation d'une série temporelle comme la superposition du signal et du bruit n'est pas intrinsèque mais dépend du but poursuivi par l'étude, des processus qui sont étudiés, des échelles de temps considérées comme utiles.

L'utilisation de la moyenne glissante comme méthode de filtrage a déjà été brièvement évoquée dans le premier chapitre. Plus généralement, on peut caractériser le processus de filtrage (linéaire) par

$$f_w(t) = \int_{-\infty}^{\infty} f(t-\tau)w(\tau)d\tau \tag{4.33}$$

où $f(\cdot)$ désigne le signal brut, $f_w(\cdot)$ est le signal filtré et $w(\tau)$ est une fonction caractérisant le filtrage. Cette fonction $w(\tau)$ introduit une pondération des valeurs prise le signal brut pour le calcul du signal filtré. On vérifie par exemple que le choix

$$w(\tau) = \begin{cases} \dfrac{1}{T} & \text{si } |\tau| < \dfrac{T}{2} \\ 0 & \text{sinon} \end{cases} \tag{4.34}$$

conduit au calcul d'une moyenne glissante sur la période T.

En s'appuyant sur l'exemple simple de la moyenne glissante, on peut aisément déterminer quelques propriétés de la fonction w permettant de définir un filtrage approprié.

Tout d'abord, afin d'éviter que le filtrage d'un signal brut positif ne donne naissance à des valeurs négative du signal extrait, on exigera généralement que la fonction w soit positive. On pourra également exiger que la fonction $w(\tau)$ soit à support fini, *i.e.* que celle-ci s'annule en-dehors d'un intervalle borné. Ceci assure qu'un événement quelconque présent dans le signal brut n'influence le signal filtré que pendant un laps de temps fini.

Ensuite, pour que le filtrage par (4.33) corresponde au calcul d'une moyenne pondérée du signal brut, il importe que la somme des poids représentés par la fonction $w(\tau)$ soit égale à 1, soit

$$\int_{-\infty}^{\infty} w(\tau)d\tau = 1 \tag{4.35}$$

Cette condition peut aussi être obtenue en considérant le cas (très) particulier d'un signal brut constant. Dans ce cas, (4.33) devrait fournir un signal filtré rigoureusement identique au signal de départ. La condition (4.35) garantit qu'il en soit bien ainsi. On vérifie que la fonction de poids (4.34) définissant la moyenne glissante respecte bien la condition (4.35).

En général, on exigera également que $w(\tau)$ soit une fonction paire de son argument, *i.e.*

$$w(\tau) = w(-\tau) \tag{4.36}$$

Cette condition permet d'éviter que l'application du filtre n'induise un déphasage du signal original. Pour s'en rendre compte, considérons un signal brut harmonique de la forme

$$f(t) = A\cos(\omega t + \varphi) \tag{4.37}$$

Il vient

$$\begin{aligned} f_w(t) &= A\int_{-\infty}^{\infty} \cos\left[\omega(t-\tau)+\varphi\right] w(\tau) d\tau \\ &= A\cos(\omega t+\varphi)\int_{-\infty}^{\infty} w(\tau)\cos\omega\tau d\tau + A\sin(\omega t+\varphi)\int_{-\infty}^{\infty} w(\tau)\sin\omega\tau d\tau \end{aligned} \tag{4.38}$$

ce qui montre que le signal voit en général non seulement son amplitude modifiée par le filtrage mais également déphasé. La condition de symétrie (4.36) permet par contre d'assurer l'annulation de la seconde intégrale de sorte que

$$f_w(t) = A\cos(\omega t+\varphi)\int_{-\infty}^{\infty} w(\tau)\cos\omega\tau d\tau \tag{4.39}$$

L'effet du filtre sur un signal harmonique se résume donc à une modification de l'amplitude selon un rapport

$$h(\omega) = \int_{-\infty}^{\infty} w(\tau)\cos\omega\tau d\tau \tag{4.40}$$

qui dépend de la pulsation du signal initial. Puisqu'un signal quelconque peut être exprimé comme la superposition d'une infinité de signaux harmoniques par le biais d'une intégrale de Fourier, l'effet du filtre peut être entièrement décrit par la donnée de la fonction $h(\omega)$ qui est appelé le *gain du filtre.*

Dans le cas de la moyenne glissante, on calcule aisément

$$h(\omega) = \frac{1}{T}\int_{-T/2}^{T/2} \cos\omega\tau d\tau = \frac{2}{\omega T}\sin\frac{\omega T}{2} \tag{4.41}$$

dont l'allure est représentée à la figure 4.5.1. On remarque que le gain est proche de l'unité pour les plus basses fréquences ; celles-ci sont donc peut affectées par le filtre. Par contre, les fréquences les plus élevées sont fortement absorbées par le filtre. Le signal filtré est donc partiellement débarrassé de ces oscillations rapides. La principale critique qui puisse être formulée au filtre ainsi réalisé tient à l'existence d'oscillations du gain aux plus hautes fréquences. Celles-ci introduisent une sélectivité non monotone telle que les plus hautes fréquences ne sont pas nécessairement amorties plus énergiquement que certaines fréquences plus basses.

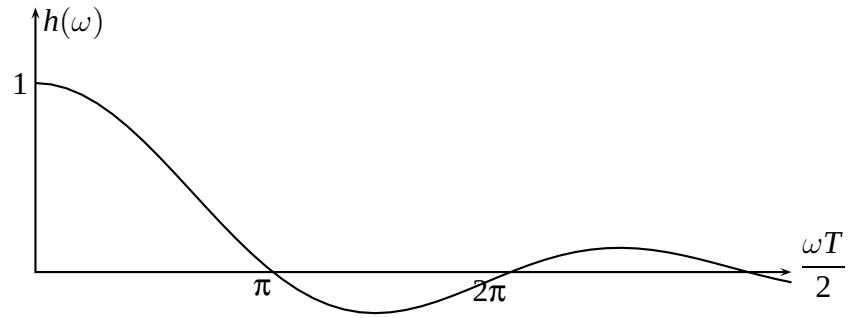


FIG. 4.4 – Gain du filtre constitué par une moyenne glissante de période T.

4.5.2 Cas discret.

Les séries temporelles dont on dispose sont généralement discrètes, *i.e.* constituées d'une liste de valeurs correspondant à des mesures supposées ici effectuées à intervalles réguliers. Bien que le formalisme continu adopté dans la section précédente n'est plus applicable, les principes sont cependant aisément transposables.

Notons x_j ($j = 1, 2, \ldots$) la suites des mesures effectuées aux temps $t_j = t_0 + j\Delta t$. Par analogie avec (4.33), le signal filtré sera désormais obtenu par

$$f_j = \sum_{k=-N}^{N} w_k x_{j+k} \tag{4.42}$$

où w_k ($k = -N, \ldots, N$) désigne les poids du filtre, supposé à support fini. Un tel filtre comportant $\ell = 2N+1$ poids est dit de longueur ℓ. L'expression (4.42), avec la condition

$$\sum_{k=-N}^{N} w_k = 1 \tag{4.43}$$

correspondant à (4.35) permet une fois encore d'interpréter le filtrage comme le calcul d'une moyenne pondérée des valeurs successives de la série temporelle où la valeur f_j est obtenue à partir des valeurs $x_{j-N}, x_{j-N+1}, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots x_{j+N}$.

Conformément à (4.36), on choisira généralement des poids symétrique, *i.e.* tels que

$$w_{-k} = -w_k, \qquad k = 1, \ldots, N \tag{4.44}$$

de façon à ne pas introduire de déphasage par application du filtre. Pour un tel filtre, la fonction de réponse fréquentielle est donnée par

$$h(\omega) = w_0 + 2\sum_{k=1}^{N} w_k \cos(k\omega\Delta t) \tag{4.45}$$

Dans le cas discret, la moyenne glissante est calculée par le biais d'un filtre symétrique de longueur $\ell = 2N+1$ dont tous les poids sont égaux, *i.e.*

$$w_{-N} = w_{-N+1} = \cdots = w_{-1} = w_0 = w_1 = \cdots = w_{N-1} = w_N = \frac{1}{\ell} \tag{4.46}$$

La réponse fréquentielle du filtre est unitaire pour $\omega = 0$ (signal constant) et décroît pour des pulsations/fréquences plus grandes. La réponse est nulle pour $\omega = \omega_T = 2\pi/(\ell\Delta t)$, *i.e.* pour une longueur d'onde correspondant à la longueur du filtre. Ainsi donc, une moyenne glissante calculée sur une période $T = \ell\Delta t$ annule exactement la composante du signal à cette période et présente une réponse croissante pour des composantes de plus grande période. Remarquons que la plus petite période pouvant être captée avec un pas d'échantillonnage Δt est $2\Delta t$. La réponse fréquentielle ne doit donc être étudiée que dans l'intervalle $\omega \in [0, \pi/\Delta t]$.

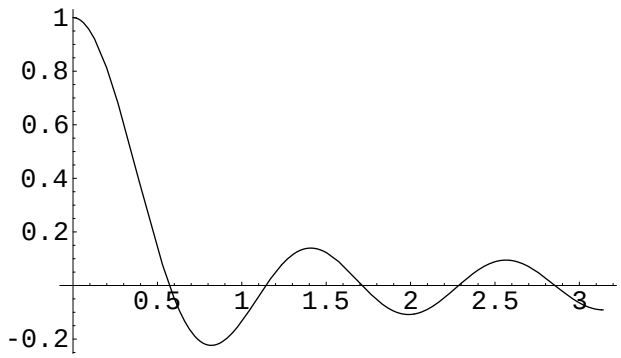


FIG. 4.5 – Réponse fréquentielle de la moyenne glissante discrète en fonction de $\omega\Delta t$ (Cas particulier $N = 5$).

Un tel filtre a évidemment l'avantage de la simplicité. Cependant, il souffre du même problème que sa version continue. Pour des signaux de période plus courte que la période T de calcul de la moyenne, *i.e. pour des pulsations supérieures à* ω_T, la réponse fréquentielle oscille autour d'une valeur nulle, ce qui peut compliquer l'interprétation de fluctuations dans la série filtrée.

4.5.3 Filtres binomial et gaussien

Pour éliminer efficacement les fluctuations aux hautes fréquences, un filtre devrait idéalement présenté une réponse fréquentielle proche de l'unité aux basses fréquences, décroissant vers zéro à une certaine fréquence de coupure et rester approximativement nulle aux fréquences plus élevées. Cette propriété peut être obtenue en utilisant des poids dont la valeur décroît progressivement à partir du poids central w_0. Les filtres binomial et gaussien possèdent cette propriété.

Les poids du filtre binomial sont choisis proportionnellement aux coefficients binomiaux. Pour un filtre de longueur $\ell = 2N+1$, on a

$$w_k = \frac{1}{2^{2N}} \frac{(2N)!}{(N-k)!(N+k)!}, \qquad k = -N, -N+1, \ldots, -1, 0, 1, \ldots, N \tag{4.47}$$

Pour réaliser un filtre binomial avec une réponse fréquentielle de 0.50 à une période T donnée, on choisit N comme l'entier le plus proche de

$$(T/\Delta t)^2/12 \tag{4.48}$$

Ainsi, pour filtrer une série de mesures annuelles avec un amortissement de 50 % de la réponse à 10 ans, on choisira $N = 6$. Les coefficients correspondants sont donc donnés par

0.00024411 0.00292969 0.0161133 0.0537109 0.12085 0.193359

0.225586

0.193359 0.12085 0.0537109 0.0161133 0.00292969 0.00024411

Pour éviter d'avoir à tenir compte de poids trop petits, on peut négliger ceux dont la valeur est inférieure à, par exemple, 5 % du poids central. Il importe cependant alors de renormaliser les poids (en divisant par la somme des poids significatifs retenus) pour que leur somme soit égale à 1. Dans le cas précédent, on peut ainsi se ramener à une filtre de longueur 9 défini par

$$0.0162 \quad 0.0541 \quad 0.1216 \quad 0.1946 \quad 0.2270 \quad 0.1946 \quad 0.1216 \quad 0.0541 \quad 0.0162 \tag{4.49}$$

Lorsque la longueur du filtre binomial devient importante, les poids présentent une distribution proche de celle d'une gaussienne. Un filtre aux propriétés semblables à celles du filtre bionomial peut donc être obtenu en déterminant les poids directement comme les ordonnées d'une distribution normale soit

$$w_k = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{k^2\Delta t^2}{2\sigma^2}\right] \qquad k = -N, -N+1, \ldots, -1, 0, 1, \ldots, N \tag{4.50}$$

L'écart-type de la distribution normale permettant un amortissement de 50 % pour une période T est donné par

$$\sigma = \frac{T}{6} \tag{4.51}$$

Ici encore, afin d'éviter de travailler avec un filtre trop long et des poids trop petits, on exclut les poids dont la valeur est inférieure à 5% du poids maximum et on renormalise les poids pour que leur somme soit égale à 1. Ainsi, dans le cas évoqué plus haut du filtrage d'une série de mesures annuelles avec un amortissement de 50 % de la réponse à 10 ans, on prendra $\sigma = 1.666$ années et les poids sont donnés par (en se limitant provisoirement à un filtre de longueur 13 comme précédemment)

0.000367141 0.00265911 0.0134367 0.0473701 0.116512 0.199935

0.239365

0.199935 0.116512 0.0473701 0.0134367 0.00265911 0.000367141

En se limitant aux poids les plus significatifs, en renormalisant les poids et en arrondissant, on obtient

$$0.0135 \quad 0.0477 \quad 0.1172 \quad 0.2012 \quad 0.2408 \quad 0.2012 \quad 0.1172 \quad 0.0477 \quad 0.0135 \tag{4.52}$$

La figure 4.6 compare les réponses fréquentielles des filtres binomial et gaussien correspondant au cas particuliers (4.49) et (4.52). On remarque que la réponse est effectivement de (approximativement) 50 % pour une pulsation $\omega\Delta t = 2\pi/10 \approx 0.63$.

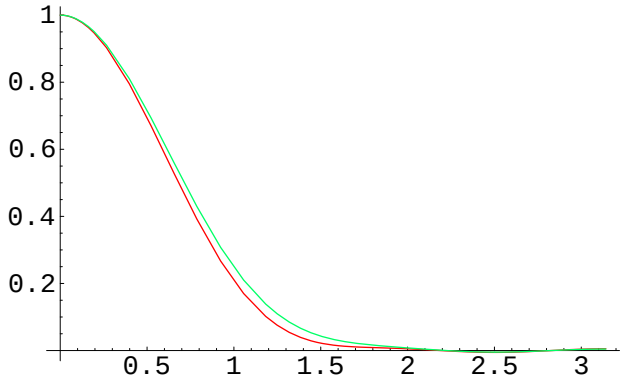


FIG. 4.6 – Réponse fréquentielle des filtres binomial (en rouge) et gaussien (en vert) en fonction de $\omega\Delta t$ dans le cas du filtrage de valeurs annuelles avec une amortissement à 50 % de la composante de période égale à 10 ans (Cas particulier (4.49) et (4.52)).

4.5.4 Fenêtre de Hamming

Comme le montre la figure 4.6, les filtres binomial et gaussien présentent bien la caractéristique recherchée de décroissance de la réponse en fonction de la fréquence et absorbent quasi complètement les signaux aux hautes fréquences. Ces filtres sont cependant perfectibles puisqu'ils affectent de façon progressive les signaux de fréquences moyennes. Un filtre idéal devrait avoir une réponse fréquentielle unitaire pour tous les signaux de fréquence inférieure à une certaine fréquence de coupure et rigoureusement nulle pour toutes les composantes de fréquence supérieure. Théoriquement, un tel filtre peut être obtenu en décomposant formellement le signal brut en ses différentes composantes de Fourier et en supprimant celles dont la fréquence est supérieure à la fréquence de coupure choisie.

Cette procédure peut être formalisée à partir du concept de transformée de Fourier et de l'expression générale (4.33) du filtrage. En termes mathématiques, on dit que le signal filtré obtenu par (4.33) est la convolution du signal original $f(t)$ avec le filtre $w(t)$.

Calculons maintenant la transformée de Fourier du signal filtré f_w. On a

$$\begin{aligned}\tilde{f}_w(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(t-\tau) w(\tau) d\tau \right\} \mathrm{e}^{-i\omega t}\, dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(t-\tau)\, \mathrm{e}^{-i\omega t}\, dt \right\} w(\tau) d\tau\end{aligned} \tag{4.53}$$

Si on pose $t' = t - \tau$ dans la seconde intégrale, on obtient

$$\begin{aligned}\tilde{f}_w(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(t')\, \mathrm{e}^{-i\omega t'}\, dt' \right\} w(\tau)\, \mathrm{e}^{-i\omega\tau} d\tau \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t')\, \mathrm{e}^{-i\omega t'}\, dt' \int_{-\infty}^{\infty} w(\tau)\, \mathrm{e}^{-i\omega\tau} d\tau \\ &= \sqrt{2\pi}\, \tilde{f}(\omega) \tilde{w}(\omega)\end{aligned} \tag{4.54}$$

Cette expression montre que la composante de pulsation ω dans le signal filtré est simplement obtenue par multiplication de la composante de même pulsation dans le signal brut multipliée par la composante correspondante du filtre. L'effet du filtre w sur un signal quelconque est donc parfaitement décrit dans le domaine fréquentiel par la donnée de sa transformée de Fourier $\tilde{w}$ qui apparaît, au facteur $\sqrt{2\pi}$ près, comme le facteur d'amortissement dont est affectée chaque composante du signal.

La fonction $w(t)$ correspondant à une filtre idéal supprimant les composantes du signal de fréquence supérieure à une fréquence de coupure ω_c donnée est donc décrite par

$$\tilde{w}_{ideal}(\omega) = \frac{1}{\sqrt{2\pi}} \begin{cases} 1 & \text{si } |\omega| < \omega_c \\ 0 & \text{si } |\omega| \geq \omega_c \end{cases} \tag{4.55}$$

et peut être formellement obtenue dans le domaine temporel en inversant la transformée de Fourier par (4.17), soit

$$\begin{aligned} w_{ideal}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{w}_{ideal}(\omega)\, \mathrm{e}^{i\omega t}\, d\omega \\ &= \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \mathrm{e}^{i\omega t}\, d\omega \\ &= \frac{\sin \omega_c t}{\pi t} \end{aligned} \tag{4.56}$$

La même procédure peut être appliquée dans le cas discret. Celle-ci conduit à choisir les poids du filtre en échantillonnant la distribution continue $w_{ideal}(t)$ aux instants correspondants aux points de support de la série temporelle. On prendra donc

$$w_{ideal,k} = w_{ideal}(k\Delta t) = \frac{\sin \omega_c k\Delta t}{\pi k \Delta t}, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{4.57}$$

Pratiquement, la construction d'un filtre sur base de (4.56) ou (4.57) est cependant irréalisable car $w(t)$ n'a pas un support borné ; le signal filtré en un instant donné dépend

théoriquement du signal à tous les instants passés et ultérieurs. Pour obtenir un filtre discret utilisable pratiquement, on peut décider de tronquer le filtre en ignorant les valeurs de $w_{ideal,k}$ pour des k trop grands, *i.e.* de restreindre les poids à une certaine fenêtre choisie de façon appropriée. Ceci conduit malheureusement à de nouvelles difficultés si la largeur de la fenêtre choisie est insuffisante et s'accompagne de l'annulation brutale des poids du filtre. On préférera donc affecter les poids successifs du filtre ideal (4.57) d'un facteur décroissant au fur et à mesure que l'on s'écarte du poids central $w_{ideal,0}$ et assurant l'amortissement progressif de la suite des poids.

Pour générer un filtre discret de longueur $\ell = 2N+1$, la fenêtre de Hamming définie par

$$h_k = 0.54 + 0.46\cos\frac{\pi k}{N} \qquad k = 0, \pm 1, \ldots, \pm N \tag{4.58}$$

est parmi les plus utilisées dans ce cadre. Dès que la longueur du filtre est choisie, les poids réels sont donc calculés selon

$$w_{ideal,k} \cdot h_k \qquad k = 0, \pm 1, \ldots, \pm N \tag{4.59}$$

puis normalisés pour que leur somme soit égale à l'unité. Le résultat est un filtre réalisable qui constitue la meilleure approximation du filtre idéal pour la longueur souhaitée. Plus le filtre est long, plus le comportement réel est proche du comportement idéal.

Il faut noter que le filtre correspondant peut comporter des poids négatifs, ce qui peut parfois être surprenant.

À titre d'exemple, si on considère une série de valeurs annuelles dont on désire supprimer les composantes de périodes inférieure à 10 ans au moyen d'un filtre de longueur 9, on calculera successivement

$$w_{ideal,k} = \frac{1}{\pi k}\sin\frac{2\pi k}{10}, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{4.60}$$

$$h_k = 0.54 + 0.46\cos\frac{\pi k}{4} \qquad k = 0, \pm 1, \ldots, \pm 4 \tag{4.61}$$

k	$w_{ideal,k}$	h_k	$w_{ideal,k} \cdot h_k$	w_k
0	0.2	1	0.2	0.271
$\pm$ 1	0.187	0.865	0.162	0.219
$\pm$ 2	0.151	0.54	0.082	0.111
$\pm$ 3	0.101	0.215	0.022	0.029
$\pm$ 4	0.047	0.08	0.004	0.005

EXEMPLE 4.3 Considérons à titre d'exemple, la série de données représentée à la figure 4.7. Cette série est composées de 128 valeurs échantillonnées avec un Δt constant. L'analyse de Fourier de ces données révèle la présence de composantes dont les périodes sont approximativement égales à 20-30 Δt, 5 Δt et 3Δt.

Dans le but de supprimer les composantes dont les fréquences sont égales à 3-5 Δt, on peut appliquer une moyenne glissante calculée sur 9 valeurs[6]. Cette procédure permet d'éliminer la plus grande partie des oscillations aux hautes fréquences. Le filtrage n'est cependant pas parfait : de petites fluctuations apparemment étrangères au signal principal sont toujours présentes (Figure 4.9).

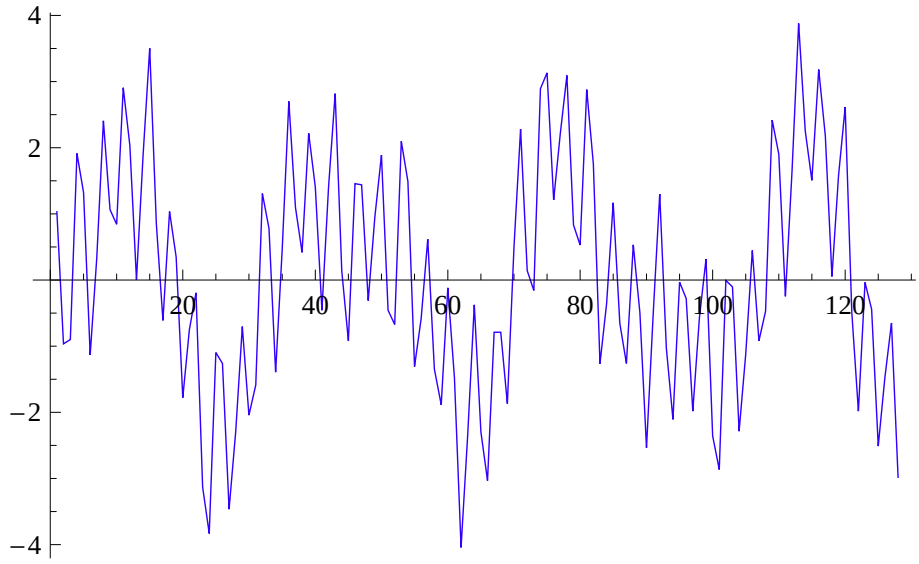


FIG. 4.7 – Série de données brutes (longueur = 128)

[6]Pratiquement, l'application d'un filtre discret de longueur $\ell = 2N+1$ n'est possible en les N premiers et N derniers instants que si on prolonge artificiellement la série de données. Ceci peut être fait en remplaçant les données par leur moyenne ou en supposant que l'enregistrement est périodique.

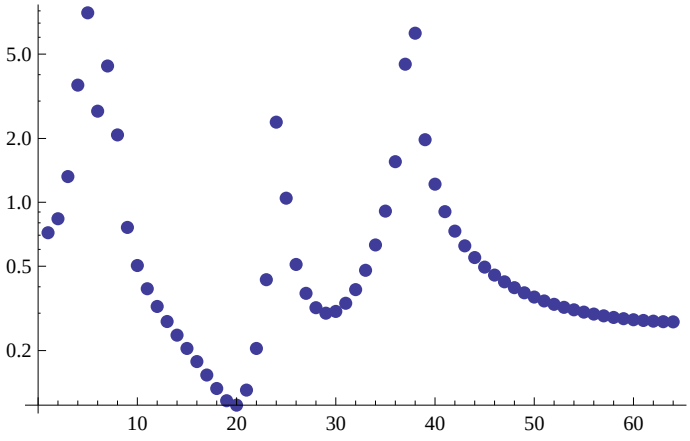


FIG. 4.8 – Transformée de Fourier des données de la figure 4.7

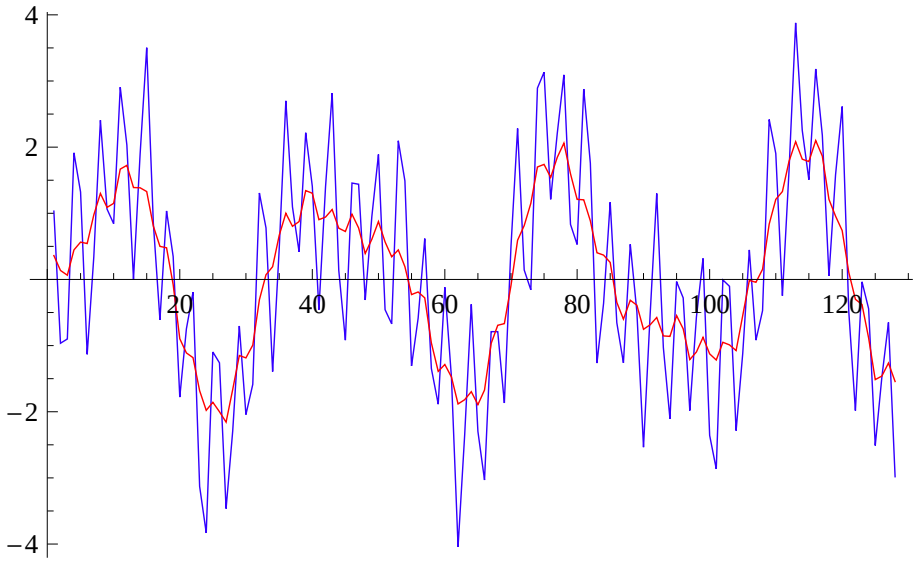


FIG. 4.9 – Moyenne glissante (largeur = 9) appliquée aux donnée de 4.7

Afin d'obtenir un filtrage plus efficace des hautes fréquences, on réalise le filtrage par un filtre gaussien et un filtre idéal avec une fenêtre de Hamming de longueur 7. Dans les deux cas, la période de coupure est choisie égale à $10\Delta t$. Les résultats obtenus avec des deux filtres sont très semblables (Figure 4.10) et éliminent parfaitement les composantes non désirées.

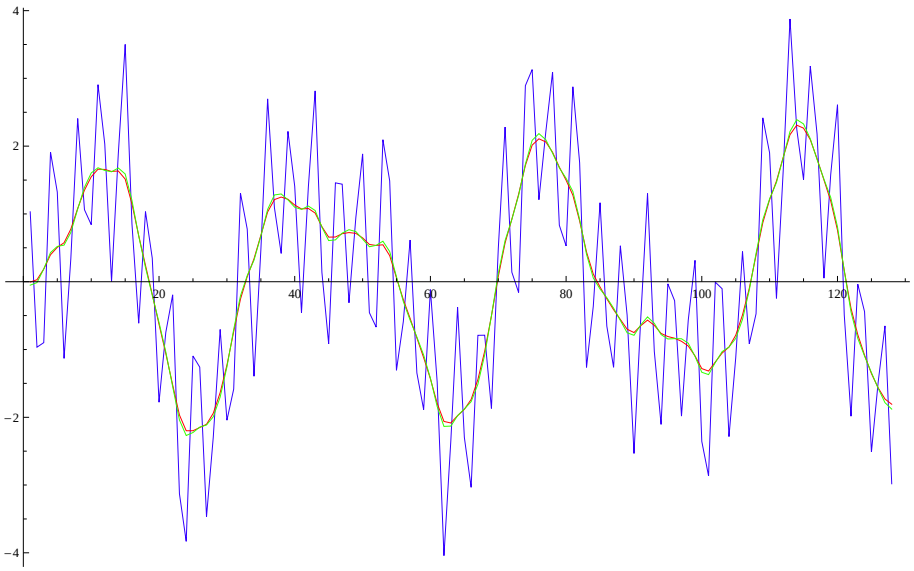


FIG. 4.10 – Filtrage gaussien (en rouge) et filtre idéal avec fenêtre de Hamming (en vert) appliqués à 4.7

L'efficacité du filtrage peut être examinée en calculant la transformée de Fourier discrète des différents signaux filtrés. On vérifie sur la figure 4.11 que les filtres appliqués induisent bien un amortissement des signaux aux hautes fréquences et affectent peu ceux de plus basses fréquences. Les filtres gaussien et idéal donnent des résultats pratiquement identiques. La figure confirme la présence significative de signaux de hautes fréquences dans la série obtenue par application de la moyenne glissante.

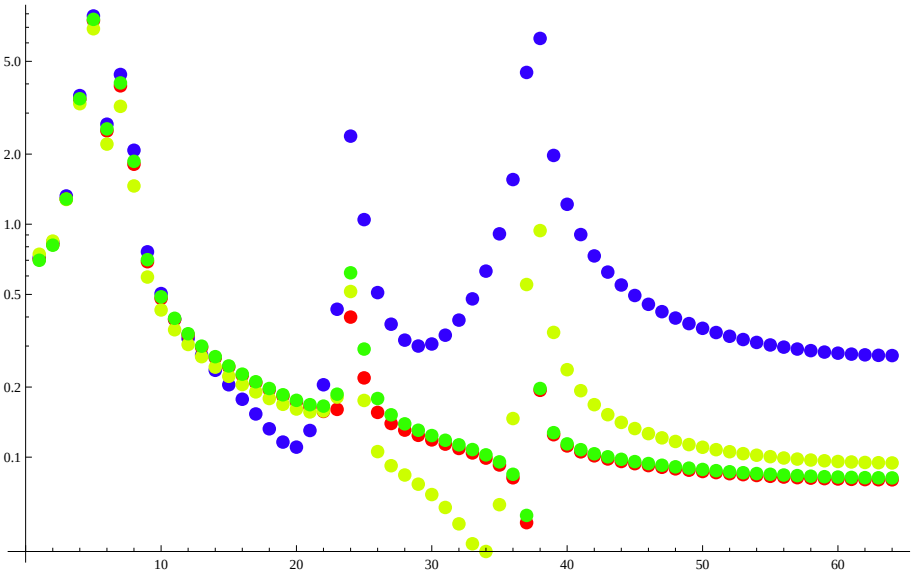


FIG. 4.11 – Transformée de Fourier du signal brut (en bleu), de la moyenne glissante (en jaune), du filtre gaussien (en rouge) et filtre idéal avec fenêtre de Hamming (en vert) appliqués à 4.7

◇

4.6 *Detrending.*

Un signal monotone n'est pas bien décrit par sa transformée de Fourier. En effet, celle-ci fait apparaître une multitude de composantes harmoniques qui couvrent une large gamme de fréquences (Figure 4.12). Dés lors, afin de ne pas polluer l'analyse par une éventuelle tendance présente dans les données analysées, il est souhaitable de soustraire cette tendance aux données avant de réaliser l'analyse de Fourier. La même procédure doit aussi être réalisée pour appliquer un grand nombre de méthodes statistiques qui supposent la stationnarité des grandeurs étudiées. Il convient alors de supprimer les variations à long terme de la moyenne, voire de la variance.

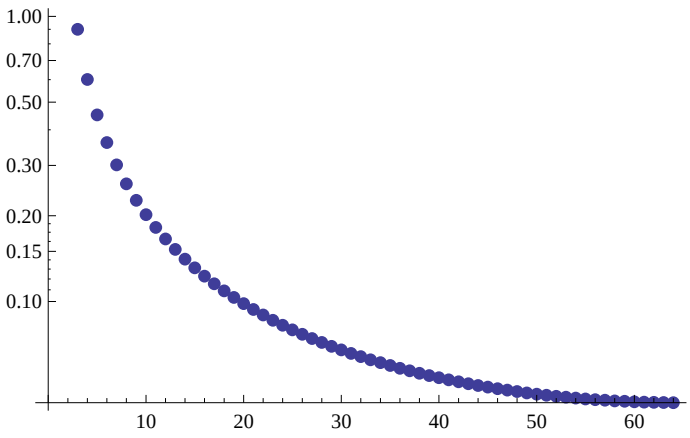


FIG. 4.12 – Transformée de Fourier du signal $t/N\Delta t$ (N=128). Le signal est formé d'un multitude de composantes dont les amplitudes décroissent avec la fréquence.

L'identification et la définition d'une tendance dans une série de données dépendent du propos de l'étude. Ainsi, ce qui apparaît comme une tendance dans une série temporelle couvrant une courte période pourra n'être que la manifestation d'une fluctuation de très basse fréquence pouvant être mise en évidence au travers de plus longues séries temporelles. Le contexte physique pourra aussi aider à l'interprétation de la nature des variations observées. En pratique, on considère généralement comme une tendance toute variation dont la période est supérieure ou égale à deux fois la longueur de la série de données étudiées.

4.6.1 Dérivation.

Une série temporelle x_j qui n'est pas stationnaire en moyenne, dont la moyenne évolue, peut être rendue stationnaire par simple dérivation, soit en remplaçant la série d'origine par

$$x'_j = x_j - x_{j-1} \tag{4.62}$$

Si la moyenne évolue de façon non constante, on pourra calculer la dérivée seconde, *i.e.* la dérivée discrète de la dérivée première.

La dérivation peut être très efficace pour atténuer les variations aux plus basses fréquences de la série temporelle. Elle peut par contre conduire à donner un poids trop grand aux variations aux hautes fréquences. Cette technique, simple, doit donc être utilisée avec prudence.

4.6.2 Filtrage.

Dans le cas où la tendance étudiée est raisonnablement décrite par les données disponibles, *i.e.* les techniques de filtrage introduites plus haut peuvent être appliquées

pour identifier et supprimer une éventuelle tendance. Dans ce cas, l'application d'un filtre dont la fréquence de coupure est très basse permet de mettre en évidence la tendance et, par différence, cette tendance peut ensuite être retirée de la série temporelle initiale.

4.6.3 Ajustement de courbe.

La tendance présente dans une série temporelle peut être mise en évidence en ajustant une courbe aux données disponibles. Le cas le plus fréquent est celui où une simple loi linéaire

$$x_j = b_0 + b_1 j + \varepsilon_j \tag{4.63}$$

est ajustée aux données disponibles en utilisant les techniques de régression linéaire.

D'autres courbes peuvent également être ajustées. La loi linéaire n'a d'autre mérite que celui de la simplicité. Seule une connaissance des processus responsables de la tendance peut permettre de justifier le choix d'une forme analytique plutôt qu'une autre.

4.6.4 Lissage spline.

Le lissage spline constitue une alternative à l'ajustement global d'une fonction unique représentant la tendance sur toute la durée de la série temporelle. Dans le cas du lissage spline cubique, la tendance est estimée par le biais d'une fonction $s(t)$ définie par morceaux tels que

- la fonction $s(t)$ se réduit à un polynôme d'ordre 3 sur l'intervalle de temps correspondant à chaque triplets de valeurs consécutives de la série temporelle ;
- les dérivées première et secondes sont continues en chaque point.

Les données de la série temporelle sont supposées représentatives d'une fonction $g(t)$ suffisamment régulière telle que

$$x_j = g(t_j) + \varepsilon_j \tag{4.64}$$

où ε_j désigne l'écart entre la valeur observée et la valeur prédite par g. À partir de d'une estimation δx_j de l'incertitude sur les données, le problème est de reconstruire la fonction $g(t)$. La spline de lissage cubique correspond au minimum de

$$p \sum_{j=0}^{N-1} \left[\frac{x_j - s(x_j)}{\delta x_j} \right]^2 + (1-p) \int_{t_0}^{t_{N-1}} \left[s''(t) \right]^2 dt \tag{4.65}$$

Le paramètre $p \in [0,1]$ introduit dans cette expression est utilisé pour pondérer l'importance des deux termes de cette expression, *i.e.* la contrainte portant sur l'interpolation des données et l'objectif de minimisation de la courbure totale. Pour une $p = 0$ la procédure est équivalente à l'ajustement d'une droite à l'ensemble des données. Pour $p = 1$, elle conduit à une interpolation cubique classique passant exactement par chaque point. Pour des valeurs intermédiaires de p, la fonction spline a un comportement mixte.

On peut montrer que la spline de lissage cubique correspond à une réponse fréquentielle donnée par

$$u(\omega) = \left[1 + 12\frac{1-p}{p}\frac{(\cos\omega - 1)^2}{(\cos\omega + 2)}\right]^{-1} \tag{4.66}$$

Cette réponse est normalement plus élevée aux basses fréquences, correspondant au souci de mettre en évidence ces composantes. Cette expression peut être utilisée pour choisir le paramètre p correspondant à un amortissement de 50% à une fréquence donnée.

Chapitre 5

Modélisation dynamique à une équation.

5.1 Introduction.

Un grand nombre de systèmes peuvent être décrits par une équation différentielle exprimant le bilan d'une grandeur caractéristique de l'état de ce système. Ainsi la masse totale M d'un constituant quelconque présente dans une région particulière de l'espace évolue en fonction du temps selon une loi du type

$$\frac{dM}{dt} = \text{Production - Destruction + Echange} \tag{5.1}$$

où les trois termes du membre de droite correspondent effectivement aux taux de production, de destruction et d'échange du constituant considéré. En général, on adopte comme convention de considérer comme positif tout apport au système tandis que les échanges avec le monde extérieur sont négatifs s'ils induisent une perte pour le système. De même la dynamique d'une population d'une espèce particulière mesurée par N (nombre d'individus ou biomasse exprimée en unités appropriées) suit une loi du type

$$\frac{dN}{dt} = \text{Natalité - Mortalité + Migration/Transport} \tag{5.2}$$

où les trois termes du membre de droite représentent respectivement les taux de natalité, de mortalité et de migration ou de transport (apports extérieurs).

Dans les cas les plus simples, les termes de production/destruction et d'échange (resp. de natalité/mortalité, migration/transport) peuvent s'exprimer directement en fonction de M (resp. de N) de sorte que l'équation de bilan permet à elle seule de déterminer l'évolution temporelle $M(t)$ (resp. $N(t)$). Dans d'autres cas, la paramétrisation des taux de variation en fonction de M (resp. de N) cache une modélisation des effets combinés d'une multitude de processus qui ne peuvent être pris en compte explicitement (ou que l'on ne désire pas prendre en compte explicitement).

5.2 Modèles différentiels.

5.2.1 Modèles malthusien et logistique.

Le modèle le plus simple de la forme (5.2) ignore la migration et le transport et suppose que la natalité et la mortalité sont proportionnelles à N. Ce modèle, dû à Malthus (1798), s'écrit donc

$$\frac{dN}{dt} = bN - dN \tag{5.3}$$

où b et d sont des constantes positives. Si la population initiale est N_0, alors

$$N(t) = N_0\, \mathrm{e}^{(b-d)t} \tag{5.4}$$

La population croît exponentiellement si $b > d$, *i.e.* si le taux de croissance net $r = b - d$ est positif, s'éteint rapidement si $b < d$ et demeure égale à sa valeur initiale si $b = d$.

Bien que le modèle de croissance exponentielle de Malthus s'applique à un certain nombre de systèmes biologiques pendant un temps limité de leur évolution (*e.g.* la croissance de la population mondiale entre le XVIIème et le XXIème siècles), ce modèle est irréaliste pour établir des prévisions à plus ou moins long terme. Une régulation de la croissance exponentielle doit généralement être prise en compte. Une telle limitation est incluse dans le modèle logistique (Verhulst, 1845)

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{5.5}$$

où r et K sont des constantes positives. Dans ce modèle, le taux de croissance spécifique $r(1 - N/K)$ est une fonction décroissante de la population pour refléter la disponibilité des ressources en quantité limitée ; lorsque la population N est grande, les ressources deviennent insuffisantes et le taux de croissance est réduit.

Si $N(0) = N_0$, la solution de (5.5) est aisément obtenue en profitant de la structure particulière de l'équation différentielle. Celle-ci est en effet à 'variables séparables'. On a successivement

$$\int_{N_0}^{N} \frac{dN'}{N'\left(1 - \frac{N'}{K}\right)} = \int_0^t r\, dt' \tag{5.6}$$

$$\int_{N_0}^{N} \left[\frac{1 - \frac{N'}{K}}{N'\left(1 - \frac{N'}{K}\right)} + \frac{\frac{N'}{K}}{N'\left(1 - \frac{N'}{K}\right)}\right] dN' = \ln\frac{N}{N_0} - \ln\frac{1 - N/K}{1 - N_0/K} = rt \tag{5.7}$$

$$N(K - N_0) = \mathrm{e}^{rt} N_0 (K - N) \tag{5.8}$$

$$N(t) = \frac{N_0 K\ \mathrm{e}^{rt}}{K + N_0(\mathrm{e}^{rt} - 1)} \tag{5.9}$$

Cette solution est représentée à la figure 5.1 pour différentes valeurs de N_0. Quelle que soit la condition initiale $N_0 \neq 0$, la population tend vers la population d'équilibre K lorsque $t \to +\infty$.

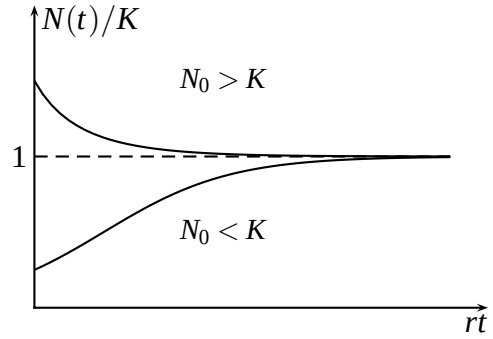


FIG. 5.1

La constante K représente donc la *capacité portante* du système ('carrying capacity'). La constante r représente le taux de croissance spécifique pour de faibles valeurs de N. Elle représente également la vitesse à laquelle la solution $N(t)$ tend vers K ($1/r$ est le temps caractéristique de la réponse).

Le comportement de la solution $N(t)$ est qualitativement indépendant des constantes r et K du problème. Ceci pouvait également se déduire de la mise sous forme adimensionnelle de l'équation (5.5). En effet, posant

$$t' = rt, \qquad N' = \frac{N}{K} \tag{5.10}$$

l'équation (5.5) s'écrit

$$\frac{dN'}{dt'} = N'(1-N') \tag{5.11}$$

dont la solution ne dépend d'aucun paramètre (sauf la condition initiale). Les variations du taux de croissance spécifique induisent une accélération ou un ralentissement de la réponse correspondant à une dilatation ou une contraction linéaire du temps. De même, les variations de K modifient uniquement l'amplitude de la réponse.

5.2.2 Équilibre et stabilité.

Le comportement de la solution (5.9) de (5.5) pouvait également être déduit de l'examen de cette équation, sans passer par sa résolution explicite. En effet, l'équation (5.5) montre que

$$\frac{dN}{dt} > 0$$

tant que $N < K (N \neq 0)$, *i.e.* la population croît monotonément tant que $N < K$. Inversement

$$\frac{dN}{dt} < 0$$

si la population est supérieure à la capacité du système. L'équilibre, correspondant à

$$\frac{dN}{dt} = 0,$$

n'est possible que si $N = K$ (ou $N = 0$).

Plus généralement, considérons une population décrite par une équation différentielle

$$\frac{dN}{dt} = f(N) \tag{5.12}$$

où $f(N)$ est une fonction (non linéaire) connue de N. Le système est dit 'en équilibre' lorsque la population reste constante au cours du temps.

Ceci n'est manifestement possible que pour une population N^* correspondant à un zéro de la fonction f, *i.e.*

$$f(N^*) = 0 \tag{5.13}$$

Les configurations d'équilibre d'un système peuvent être classées selon leur stabilité, *i.e.* selon le type de réponse du système lorsque la configuration stable est perturbée. Pour étudier cette réponse, écrivons $N(t)$ sous la forme

$$N(t) = N^* + \varepsilon(t) \tag{5.14}$$

La fonction $\varepsilon(t)$ représente la perturbation de l'équilibre N^*. Celle-ci vérifie l'équation différentielle

$$\frac{d\varepsilon}{dt} = f(N^* + \varepsilon) \tag{5.15}$$

En supposant que la perturbation $\varepsilon(t)$ est faible, le second membre de (5.15) peut être approché en linéarisant les variations de f au voisinage de N^*. Par le théorème de Taylor, on obtient, en tenant compte de (5.13),

$$\frac{d\varepsilon}{dt} \sim f'(N^*)\varepsilon \Leftrightarrow \varepsilon(t) = \varepsilon(0)e^{f'(N^*)t} \tag{5.16}$$

Ainsi, si $f'(N^*) < 0$, toute perturbation est amortie exponentiellement et le système retourne à sa position d'équilibre (en un temps infini, il est vrai) ; l'équilibre est dit stable. Si, par contre, $f'(N^*) > 0$, les perturbations croissent exponentiellement ; l'équilibre est dit instable. [1]

[1] Si $f'(N^*) = 0$, l'équation (5.16) devient $\dot{\varepsilon} = 0$, *i.e.* les perturbations ne sont ni amorties ni croissantes. L'équilibre est dit 'marginalement stable au sens de l'analyse linéaire'. En réalité, l'étude de la stabilité ne peut être raisonnablement menée à partir de la linéarisation (5.15). Il convient de poursuivre le développement de Taylor de f jusqu'au premier terme non nul. Soit k l'ordre de la première dérivée non

Dans ce cas, $\varepsilon(t)$ grandit au cours du temps et la linéarisation (5.16) de (5.15) introduite pour étudier la stabilité cesse d'être valable. En général, les processus ignorés par la linéarisation limitent la croissance de la perturbation. L'équilibre reste cependant réputé instable.

Les positions d'équilibre instable doivent être considérées comme des curiosités mathématiques. En effet, bien qu'une solution du type $N(t) = N^*$ soit prédite mathématiquement en un tel point, cette solution n'a pratiquement aucune chance d'être observée (ou même calculée) en pratique puisque la moindre perturbation (ou la moindre erreur d'arrondi) induite, par exemple, par des processus négligés dans la modélisation, est de nature à faire s'écarter le système de la solution d'équilibre instable prédite théoriquement.

L'analyse de stabilité constitue un outil extrêmement utile pour examiner l'applicabilité d'un modèle mathématique à un système écologique donné. En effet, si l'observation révèle l'existence d'une configuration d'équilibre (stable) du système et que le modèle mathématique n'admet pas de solution d'équilibre stable, la structure du modèle est clairement inadaptée.

L'analyse linéaire de la stabilité possède des limites : elle ne nous renseigne correctement sur le comportement du système que pour des perturbations de petite amplitude. Elle ne nous donne par contre aucune information sur la réponse du système à des perturbations d'amplitude finie. Supposons par exemple que $f(N)$ présente l'allure illustrée à la figure 5.2

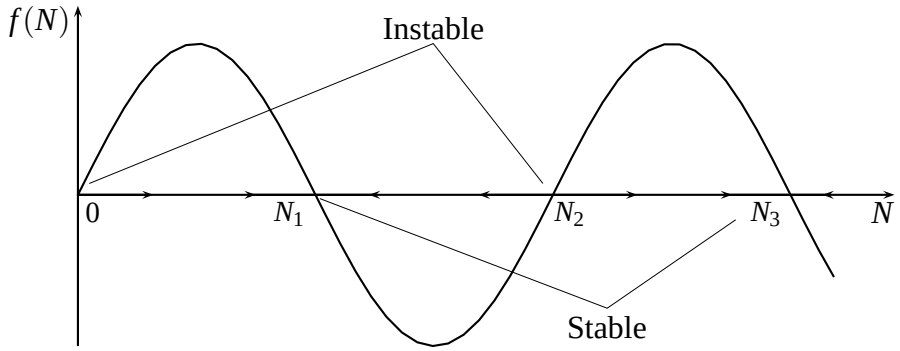


FIG. 5.2

On constate immédiatement que le système correspondant possède 4 positions

nulle de f en N^*, il vient

$$\frac{d\varepsilon}{dt} \sim \frac{f^{(k)}(N^*)}{k!}\varepsilon^k \qquad \Rightarrow \qquad \varepsilon(t) = \left[(k-1)\left(\frac{\varepsilon(0)^{1-k}}{k-1} - \frac{f^{(k)}(N^*)t}{k!}\right)\right]^{1/1-k}$$

Si $f^{(k)}(N^*) > 0$, $\varepsilon(t) \to +\infty$ et l'équilibre est instable.
Si $f^{(k)}(N^*) < 0$, $\varepsilon(t) \to 0$ et l'équilibre est stable.

d'équilibre : 0 et N_2 sont instables tandis que N_1 et N_3 sont stables au sens de l'analyse linéaire. La stabilité linéaire de N_1 peut cependant être malmenée par des perturbations de grande amplitude. Ainsi, si le système initialement en N_1 est perturbé suffisamment pour amener N dans l'intervalle $]N_2, N_3[$, son évolution ultérieure sera caractérisée par une convergence exponentielle vers N_3 et non N_1. La position $N = N_1$ est donc instable aux perturbations d'amplitude supérieure à $N_2 - N_1$.

5.2.3 Modèle de gestion des pêches et temps de recouvrement.

Un des buts de la modélisation est de permettre de définir des bases scientifiques appropriées pour la gestion des ressources et, par exemple, pour établir les quotas de pêches permettant le développement durable des espèces pêchées tout en maximisant les prises. L'exemple suivant permettra de comprendre cette problématique tout en complétant le concept de stabilité introduit précédemment.

Considérons une population dont la dynamique est décrite par un modèle de croissance logistique. Le prélèvement associé à la pêche induit un terme de perte supplémentaire qui, si l'effort de pêche est constant, peut être supposé linéaire en l'importance de la population, *i.e.*

$$\frac{dN}{dt} = r\,N\left(1 - \frac{N}{K}\right) - EN \tag{5.17}$$

Si $E < r$, la population tend vers l'équilibre stable

$$N^*(E) = K\left(1 - \frac{E}{r}\right) > 0 \tag{5.18}$$

Le prélèvement correspondant est donné par

$$P(E) = E \cdot N^*(E) = E\,K\left(1 - \frac{E}{r}\right) \tag{5.19}$$

Il est maximum pour $E = r/2$ et vaut alors $P_{\max} = rK/4$.

Se basant sur cette analyse, le gestionnaire recommandera donc qu'un effort de pêche $r/2$ soit accompli afin de maximiser les prises.

L'analyse ci-dessus suppose que l'état d'équilibre est toujours approximativement réalisé. En réalité, la constante de temps caractérisant la convergence vers cet état d'équilibre, appelée *temps de recouvrement*, est donnée par

$$T = \frac{1}{r - E} \tag{5.20}$$

Elle augmente donc avec E et, pour $E = r/2$, elle est égale à $2/r$, soit deux fois le temps caractéristique en l'absence de prélèvement par la pêche.

En pratique, l'effort de pêche ne peut être aisément quantifié. Par contre, le gestionnaire dispose de statistiques de prise permettant d'évaluer le prélèvement P. En fonction de cette grandeur, le temps de recouvrement T est donné par

$$T(P) = \frac{\frac{2}{r}}{1 \pm \sqrt{1 - \frac{P}{P_{\max}}}} \tag{5.21}$$

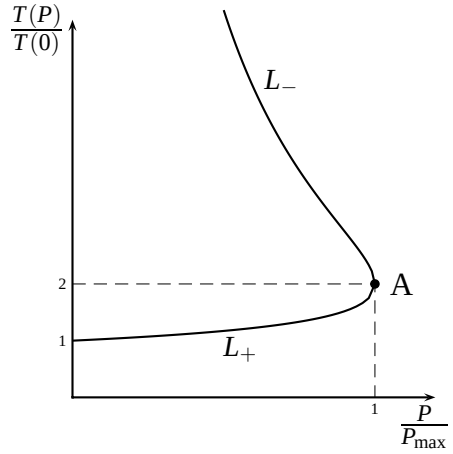


FIG. 5.3

Le graphe de cette fonction est présenté à la figure 5.3. Si l'effort de pêche E est faible, il en est de même du prélèvement P et le temps de recouvrement est proche de $1/r$. Si E augmente, le point caractéristique se déplace sur la branche L_+ de la figure 5.3. Pour $E = r/2$, le prélèvement P est égal à sa valeur maximale $P_{\max}$ et le point A est atteint. Pour une valeur supérieure de E, le prélèvement P diminue mais le temps de recouvrement augmente fortement en suivant la branche L_-.

La stratégie optimale consiste évidemment à se placer aussi proche que possible du point correspondant à $E = r/2$, mais avec $E < r/2$ pour demeurer sur la branche L_+. La difficulté réside dans le fait que, en pratique, l'effort E n'est pas connu a priori mais approché par essais et erreurs en recherchant le prélèvement P maximum. Ce faisant, on explore la région $E > r/2$ et, avec elle, la branche L_- de la figure 5.3. Si le modèle est correct, ceci peut être catastrophique puisque le temps de recouvrement de l'écosystème est alors très grand si bien qu'une réduction de l'effort de pêche peut être insuffisante pour revenir à une situation de développement stable.

5.2.4 Modèle de croissance logistique avec retard.

Le modèle de croissance logistique peut être raffiné pour tenir compte du délai induit par la période de gestation finie, le temps nécessaire pour atteindre la maturité, ...

Quantifiant ce délai par un retard unique $T > 0$, on aura, par exemple,

$$\frac{dN}{dt} = r\,N(t)\left(1 - \frac{N(t-T)}{K}\right) \tag{5.22}$$

L'introduction d'un retard dans les équations est de nature à induire un comportement oscillatoire de la solution. Qualitativement, un tel comportement peut s'expliquer en considérant la figure 5.4. Si $N = K$ à l'instant t_1 et si $N(t) < K$ pour $t_1 - T < t < t_1$, alors le membre de gauche de (5.22) est positif dans l'intervalle $[t_1, t_1 + T[$ et s'annule en $t_1 + T$. La population est alors maximale. Aux instants ultérieurs, la population décroît puisque $N(t-T) > K$ pour $t > t_1 + T$. Cette décroissance se poursuit jusqu'à l'instant $t_2 + T$ où t_2 correspond au moment où $N(t_2) = K$ (avec $t_2 > t_1$). En continuant ce raisonnement, on voit que des oscillations de période approximativement égale à $4T$ sont possibles.

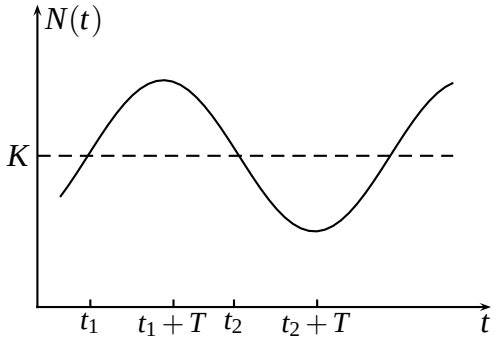


FIG. 5.4 – Solution périodique du modèle (5.22)

Étudions maintenant (5.22) plus en détail pour montrer que ces oscillations peuvent devenir instables si le retard devient important.

En utilisant les variables adimensionnelles

$$N' = \frac{N}{K}, \qquad t' = rt, \qquad T' = rT \tag{5.23}$$

on a

$$\frac{dN'}{dt'} = N'(t')[1 - N'(t' - T')] \tag{5.24}$$

Cette équation ne peut être résolue complètement à partir de la seule condition initiale $N(0)$; il est en effet nécessaire de disposer de l'évolution $N(t)$ de la population pour $-T < t < 0$. La résolution de ce type d'équation avec retard est également considérablement plus compliquée que celle d'une équation différentielle habituelle.

Les configurations d'équilibre peuvent cependant encore être étudiées comme précédemment. Les solutions de la forme $N'(t') = N'^{*} = \mathrm{C^{te}}$ vérifient

$$0 = \frac{dN'}{dt'} = N'(1 - N') \qquad \Rightarrow \qquad N' \in \{0, 1\} \tag{5.25}$$

Les configurations d'équilibre sont donc identiques à celles du cas sans retard. Leur stabilité peut être étudiée par linéarisation comme précédemment.

Au voisinage de la configuration d'équilibre $N' = 1$, on obtient

$$\frac{d\varepsilon'}{dt'} \sim \varepsilon'(t' - T') \tag{5.26}$$

où

$$\varepsilon'(t) = 1 - N'(t') \tag{5.27}$$

désigne la perturbation de l'équilibre (supposée faible, i.e. $\varepsilon' \ll 1$). Recherchons une solution de (5.26) de la forme

$$\varepsilon'(t') = C\,\mathrm{e}^{\lambda t'} \tag{5.28}$$

En substituant cette expression dans (5.26), on constate que le problème admet des solutions de la forme (5.28) pour les valeurs de λ vérifiant

$$\lambda = -\,\mathrm{e}^{-\lambda T'} \tag{5.29}$$

Les solutions réelles de (5.29) peuvent peuvent être obtenues graphiquement en interprétant les solutions de (5.29) comme le(s) point(s) d'intersection de la fonction $-\exp(\lambda T')$ avec la première bissectrice (Fig. 5.5).

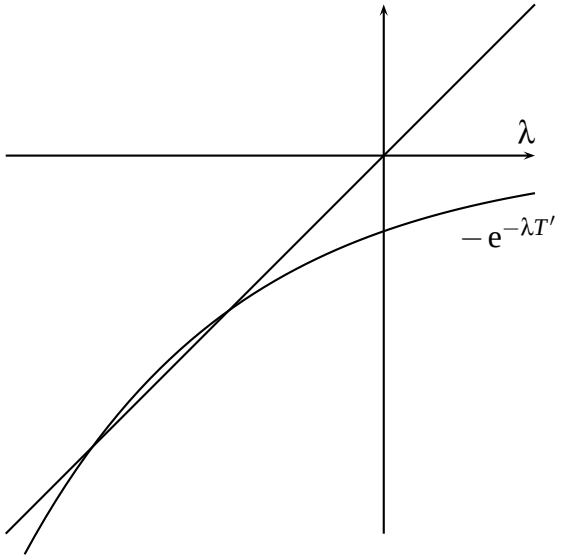


FIG. 5.5

Toutes les solutions réelles correspondent à des $\lambda < 0$. Elles donnent donc lieu à un amortissement exponentiel de la perturbation de (5.28) caractéristique d'un système asymptotiquement stable. Dans ce cas, il n'y a donc pas d'oscillations.

Pour certaines valeurs de T, l'équation (5.29) admet cependant des solutions complexes $\lambda = \sigma + i\omega$, donc oscillatoires[2]. Si σ, la partie réelle de λ, est positive, ces solutions sont instables. Pour explorer ces solutions, décomposons (5.29) en ses parties réelles et imaginaires, on a

$$\sigma = -e^{-\sigma T'} \cos \omega T', \quad \omega = e^{-\sigma T'} \sin \omega T' \tag{5.31}$$

La résolution de ce système de deux équations pour les deux inconnues (σ, ω) en fonction du paramètre T fournit les exposants $\lambda(T) = \sigma(T) + i\omega(T)$ des solutions de la forme (5.28). Pour $T = 0$, par exemple, on trouve $\sigma = -1$ et $\omega = 0$; ce qui montre la stabilité de l'équilibre $N' = 1$ dans ce cas.

Sans résoudre explicitement (5.31), on peut examiner si des solutions existent pour lesquelles $\sigma > 0$, ce qui entraînerait l'instabilité de l'équilibre. Lorsque T augmente depuis la valeur nulle, σ augmente progressivement et s'annule lorsque, en utilisant la première équation de (5.31),

$$0 = -1 \cos \omega T' \qquad \Rightarrow \qquad \omega T' = \frac{\pi}{2} + k\pi \tag{5.32}$$

Supposant $\omega > 0$ (si ω est solution de (5.31), $-\omega$ est également solution), on constate que σ s'annule pour la première fois pour $\omega T' = \pi/2$. Injectant cette expression dans la seconde équation de (5.31), on a $\omega = 1$ et donc $T' = \pi/2$.

Pour un retard $T' = \pi/2$, les perturbations de l'équilibre $N' = 1$ ne sont pas amorties $(\sigma = 0)$ mais des oscillations de pulsation (adimensionnelle) $\omega = 1$ apparaissent : la population oscille autour de sa valeur d'équilibre.

Des perturbations (5.28) instables sont possibles dès que $T' > \pi/2$, *i.e.* $T > \pi/(2r)$. L'introduction d'un retard supérieur à $\pi/(2r)$ est donc de nature à déstabiliser l'équilibre $N^* = K$. Pour T légèrement supérieur à $\pi/(2r)$, les perturbations sont caractérisées par une pulsation $\omega(T)$ proche de 1, *i.e.* elles s'accompagnent d'oscillations de période $2\pi/r$ dont l'amplitude croît exponentiellement.

La déstabilisation du système avec l'augmentation de T constitue un résultat relativement général des systèmes avec retard.

5.3 Modèles discrets.

Dans les modèles précédents, les variables sont supposées varier de façon continue au cours du temps. Parfois, on est amené à considérer des variations discontinues, soit

[2] On se souviendra que

$$e^{ix} = \cos x + i \sin x, \qquad \cos x = \frac{e^x + e^{-x}}{2}, \qquad \sin x = \frac{e^x + e^{-x}}{2i} \tag{5.30}$$

Dès lors, si $\lambda = \sigma + i\omega$, on a

$$e^{\lambda t} = e^{\sigma t}(\cos \omega t + i \sin \omega t)$$

à cause de la nature même des phénomènes étudiés (caractéristiques génétiques d'une génération à l'autre) soit par l'échantillonnage de processus continus (e.g. variations des stocks de poisson d'un hiver à l'autre). Dans ce cas, on note N_k $(k = 1, 2, \ldots)$ la variable N à l'instant $t_k = t_0 + k\Delta t$.

L'évolution de N_k est alors décrite par une *équation aux différences*, ou équation de récurrence qui, dans les cas les plus simples, s'écrit

$$N_{k+1} = f(N_k) \tag{5.33}$$

où f désigne une fonction connue. À partir de la donnée d'une condition initiale N_0 quelconque, l'application (5.33) permet de déterminer successivement $N_1, N_2, N_3, \ldots$ Cette équation constitue donc l'équivalent discret des équations différentielles utilisées dans les sections précédentes pour la modélisation des processus dynamiques.

À titre d'exemple, considérons d'abord la version discrète du modèle de croissance de Malthus :

$$N_{k+1} = r\, N_k \tag{5.34}$$

où $r > 0$ désigne le taux de reproduction net. On obtient aisément

$$N_k = r\, N_{k-1} = r(r\, N_{k-2}) = \cdots = r^k N_0 \tag{5.35}$$

Si $r > 1$, la population N_k croît exponentiellement. Celle-ci décroît exponentiellement si $r < 1$ et reste constante à sa valeur initiale pour $r = 1$. Comme le modèle continu correspondant, ce modèle est trop simple pour pouvoir être utilisé en dehors d'une période initiale de croissance d'une population pendant laquelle les ressources ne manquent pas.

5.3.1 Classification et résolution des équations aux différences

La forme la plus générale que puisse prendre une équation aux différences est

$$N_{k+1} = f(k, N_k, N_{k-1}, \ldots, N_{k-n+1}) \tag{5.36}$$

Dans ce cas, la valeur de la variable au nouvel instant N_{k+1} dépend de l'indice k et des valeurs prises aux n instants précédents $N_k, N_{k-1}, \ldots, N_{k-n+1}$. Une telle équation aux différences est dite d'ordre n. Sa résolution requiert en général la connaissance des valeurs prises par N en n instants successifs (par exemple $N_0, N_1, N_2, \ldots, N_{n-1}$).

La résolution de (5.36) est généralement impossible analytiquement lorsque f est une fonction non linéaire quelconque. Des méthodes systématiques peuvent par contre être appliquées aux équations linéaires du type

$$a_n N_{k+n} + a_{n-1} N_{k+n-1} + \cdots + a_1 N_{k+1} + a_0 N_k = g(k) \tag{5.37}$$

où $a_0, a_1, \ldots, a_n$ désignent des constantes quelconques $(a_n \neq 0)$. On montre que la solution générale de (5.37) peut s'écrire sous la forme

$$N_k = N_k^h + N_k^{part} \tag{5.38}$$

où N_k^h désigne la solution générale de l'équation homogène

$$a_n N_{k+n} + a_{n-1} N_{k+n-1} + \cdots + a_1 N_{k+1} + a_0 N_k = 0 \tag{5.39}$$

et où N_k^{part} représente une solution particulière de (5.37). La solution générale N_k s'exprime au moyen de n constantes d'intégration apparaissant dans N_k^h. Celles-ci peuvent être fixées au moyen de n conditions initiales appropriées.

Si on recherche des solutions de l'équation homogène de la forme

$$N_k = C\, z^k$$

on vérifie aisément que z doit être un zéro du polynôme caractéristique

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0$$

Comme polynôme de degré n, celui-ci possède toujours n zéros (éventuellement complexes) comptés avec leur multiplicité. Si on désigne par λ_i ($i = 1, 2, \ldots, s$) les zéros de distincts de multiplicité α_i (on a donc $\alpha_1 + \alpha_2 + \cdots + \alpha_s = n$). La solution générale de (5.39) s'écrit

$$N_k = \sum_{i=1}^{s} \mathcal{P}_i^{\alpha_i - 1}(k) \lambda_i^k \tag{5.40}$$

où $\mathcal{P}_i^{\alpha_i - 1}(k)$ désigne un polynôme de degré $\alpha_i - 1$ en k et dont les coefficients sont quelconques.

Une solution particulière de l'équation non homogène (5.37) peut être aisément déterminée dans le cas où le second membre est de la forme

$$g(k) = \mathcal{P}^p(k) \lambda^k \tag{5.41}$$

où $\mathcal{P}_{(k)}^p$ désigne un polynôme de degré $\mathcal{P}$. Il suffit en effet de rechercher une solution de la forme

$$N_k^{part} = k^\alpha \mathcal{P}_*^p(k) \lambda^k \tag{5.42}$$

où α désigne la multiplicité de λ comme zéro du polynôme caractéristique associé à l'équation homogène et où $\mathcal{P}_*^p(k)$ désigne un polynôme de degré p dont les coefficients peuvent être déterminés en substituant (5.42) dans (5.37).

À titre d'exemple, considérons le problème de multiplication des lapins proposé par Leonardo da Pisa, connu à partir du XVIII$^{\text{ème}}$ siècle sous le nom de Fibonacci. Soit donc un couple de jeunes lapins (un mâle et une femelle) abandonnés sur une île au début de l'année. Supposant que chaque couple de lapins âgés de deux mois ou plus donne naissance chaque mois à un nouveau couple de lapins, on se propose de déterminer le nombre de lapins sur l'île à la fin de l'année. On suppose également que la mortalité est nulle pendant cette année.

Notons pour ce faire N_k le nombre de couples de lapins après k mois. On a $N_0 = 1, N_0 = 1$ puisque les lapins sont trop jeunes pour se reproduire. Ensuite, le couple initial donne naissance à un nouveau couple de jeunes lapins et $N_2 = 2$. Plus généralement, à la

fin du mois k, les lapins présents sur l'île sont ceux présents au mois précédent, soit N_{k-1}, et les lapins nouveaux-nés engendrés par les couples en âge de se reproduire, c'est-à-dire ceux qui étaient présents deux mois plus tôt, *i.e.* N_{k-2}. On a donc

$$N_k = N_{k-1} + N_{k-2} \quad (k \geq 2) \tag{5.43}$$

Cette relation constitue une équation aux différences homogènes, linéaire d'ordre 2. À partir des conditions initiales $N_0 = N_1 = 1$, elle permet de déterminer successivement $N_2, N_3, \ldots$ ce qui engendre la suite de Fibonacci

$$1, 1, 2, 3, 5, 8, 13, \ldots$$

Pour obtenir une expression analytique des éléments de cette suite, on recherche les zéros du polynôme caractéristique

$$z^2 - z - 1 = 0 \qquad \Rightarrow \qquad \begin{cases} z_1 = \dfrac{1+\sqrt{5}}{2} \\ \\ z_2 = \dfrac{1-\sqrt{5}}{2} \end{cases} \tag{5.44}$$

La solution générale s'écrit donc

$$N_k = C_1 z_1^k + C_2 z_2^k \tag{5.45}$$

À partir des conditions initiales $N_0 = N_1 = 1$, on fixe la valeur des constantes d'intégration C_1 et C_2. La solution complète s'écrit finalement

$$N_k = \frac{1}{2}\left(1 + \frac{1}{\sqrt{5}}\right) z_1^k + \frac{1}{2}\left(1 - \frac{1}{\sqrt{5}}\right) z_2^k \tag{5.46}$$

On calcule aisément $N_{12} = 233$, de sorte que l'île est habitée par 466 lapins à la fin de l'année sous les hypothèses envisagées

Remarquons que, lorsque k devient grand,

$$N_k \sim \frac{1}{2}\left(1 + \frac{1}{\sqrt{5}}\right) z_1^k \tag{5.47}$$

puisque $z_1 > z_2$ et

$$\frac{N_{k+1}}{N_k} \sim z_1 = \frac{1}{\sqrt{5}} \tag{5.48}$$

Ce rapport n'est rien d'autre que le nombre d'or cher aux artistes, géomètres et philosophes de l'Antiquité et de la Renaissance.

5.3.2 Rapport avec les équations différentielles.

La classification et les techniques de résolution des équations aux différences sont en tous points semblables aux concepts correspondant dans la théorie des équations différentielles ordinaires. Ce rapport peut encore être rendu plus explicite si on remarque que les équations aux différences apparaissent naturellement lorsque l'on désire résoudre numériquement des équations différentielles.

Pour guider le raisonnement, considérons l'équation

$$u\frac{dT}{dx} = \kappa\frac{dT}{dx^2} \tag{5.49}$$

décrivant la distribution spatiale de la température soumise à l'advection à la vitesse u et à la diffusion caractérisée par le coefficient de diffusion κ. Le problème (5.49) est stationnaire et uni-dimensionnel avec $x \in [0,L]$.

La résolution numérique de (5.49) passe généralement par la discrétisation spatiale du problème. Au lieu de décrire $T(x)$ comme un champ continu, on s'intéresse seulement aux valeurs prises par la température aux noeuds x_k d'un réseau couvrant le domaine d'étude $[0,L]$. Pour simplifier, supposons les noeuds x_k régulièrement espacés avec

$$x_k = k\,\Delta x \qquad k \in \{0,1,2,\ldots,N\}$$

et $x_N = N\Delta x = L$. Parallèlement, notons T_k l'approximation discrète du champ continu $T(x)$ au point x_k.

Pour approcher les différents termes de (5.49) à partir de la seule connaissance des valeurs T_k aux noeuds, on remplace les dérivées par leurs approximations par différences finies comme à la section 1.3.2. En supposant $T(x)$ suffisamment régulière, on peut écrire

$$\left(\frac{dT}{dx}\right)(x_k) = \frac{T_{k+1}-T_k}{\Delta x} + o(\Delta x) \tag{5.50}$$

ou

$$\left(\frac{dT}{dx}\right)(x_k) = \frac{T_k-T_{k-1}}{\Delta x} + o(\Delta x) \tag{5.51}$$

ou encore

$$\left(\frac{dT}{dx}\right)(x_k) = \frac{T_{k+1}-T_{k-1}}{\Delta x} + o(\Delta x^2) \tag{5.52}$$

L'approximation centrée (5.52) de la dérivée apparaît donc la plus précise. Ceci peut être intuitivement justifié par le fait que l'approximation centrée, au contraire des deux premières approximations, ne privilégie ni les $x > x_k$, ni les $x < x_k$. Dans la discrétisation de (5.49), nous remplacerons donc la dérivée première par l'approximation centrée.

De même, la dérivée seconde présente dans le membre de droite peut être approchée au second ordre en Δx en utilisant

$$\left(\frac{d^2T}{dx^2}\right)(x_k) = \frac{T_{k+1} + T_{k-1} - 2T_k}{\Delta x^2} + o(\Delta x^2) \tag{5.53}$$

À des termes en $o(\Delta x^2)$ près, l'équation différentielle du second ordre (5.49) peut donc être approchée par l'équation aux différences du second ordre

$$u\frac{T_{k+1} - T_{k-1}}{2\Delta x} = \kappa\frac{T_{k+1} - 2T_k + T_{k-1}}{\Delta x^2} \tag{5.54}$$

soit encore

$$(P_e - 2)T_{k+1} + 4T_k - (P_c + 2)T_{k-1} = 0 \tag{5.55}$$

en introduisant le nombre (adimensionnel) de Peclet

$$P_e = \frac{u\,\Delta x}{\kappa} \tag{5.56}$$

Dans le cas où le nombre de Peclet est positif et diffère de 2, les zéros du polynôme caractéristique

$$(P_e - 2)z^2 + 4z - (P_e + 2) \tag{5.57}$$

associé à (5.55) sont donnés par

$$z_1 = 1; \quad z_2 = \frac{2 + P_e}{2 - P_e} \tag{5.58}$$

et la solution générale s'écrit

$$T_k = \alpha + \beta\left(\frac{2 + P_e}{2 - P_e}\right)^k \tag{5.59}$$

où α et β sont des constantes d'intégration à fixer en fonction des conditions aux limites du problème.

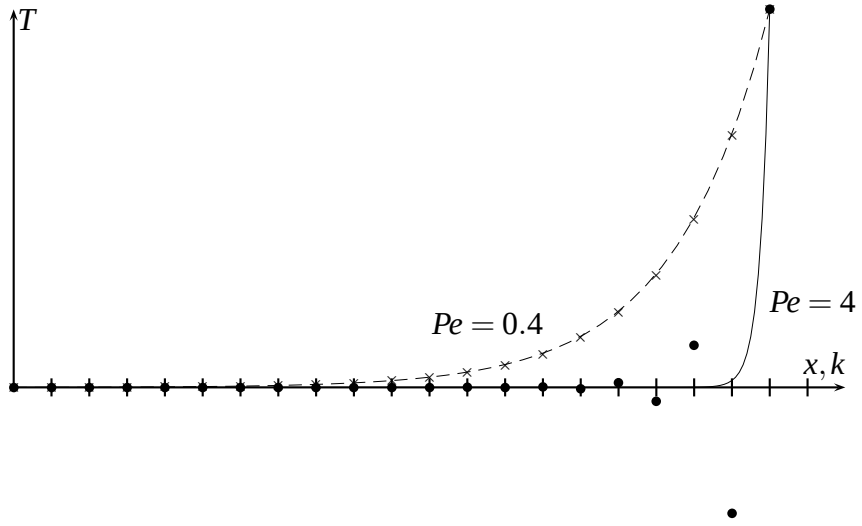


FIG. 5.6 – Solutions discrètes et continues d'un problème aux limites pour Pe=0.4 (trait pointillé et x) et Pe=4 (trait continu et •).

La solution discrète obtenue par résolution de l'équation aux différences constitue une approximation des valeurs réelles du champ continu aux noeuds du maillage. Cette approximation est d'autant meilleure que Δx (ou P_e) est petit puisque l'erreur commise en substituant l'équation aux différences à l'équation différentielle est $o(\Delta x^2)$. À partir de certaines valeurs de Δx ou de P_e, le problème discret peut cependant présenter un comportement qualitativement très différent de celui du problème continu. Ainsi, si $P_e > 2$, on remarque que z_2 devient négatif. Pour les valeurs paires et impaires de k, T_k prend des valeurs de signes différents : la solution présente un caractère oscillatoire[3] qui ne correspond pas du tout à la physique du problème continu. On dit dans ce cas que le schéma de discrétisation est instable. Pour résoudre numériquement le problème, il faut alors recourir à d'autres types d'approximation de la dérivée. Dans le problème étudié, par exemple, il suffit de remplacer l'approximation centrée de la dérivée première par une approximation décentrée (si $u > 0$) pour éviter les oscillations.

5.3.3 Analyse qualitative des systèmes discrets non linéaires

Il n'existe pas de méthode générale de résolution des équations aux différences non linéaires. Des techniques d'analyse permettent cependant d'obtenir des informations qualitatives très utiles.

[3]Remarquons que le comportement asymptotique (pour $k \to \infty$) de la solution générale (5.40) d'un problème linéaire homogène dépend du zéro λ_{max} de plus grande norme. La solution est oscillatoire si λ_{max} n'est pas un réel positif. Elle croît exponentiellement si $|\lambda| > 1$ et décroît exponentiellement si $|\lambda| < 1$. Si $|\lambda| = 1$, la solution est purement oscillatoire (ou constante) si λ_{max} est un zéro simple du polynôme caractéristique et elle est présente une croissance polynomiale sinon.

Considérons la loi de Ricker

$$N_{t+1} = f(N_t) = N_t \ \mathrm{e}^r \left(1 - \frac{N_t}{K}\right) \tag{5.60}$$

où $r, K > 0$. Cette équation prévoit une croissance exponentielle pour de faibles valeurs de N et une décroissance de taux de croissance pour N grand. Elle peut donc être vue comme équivalente à la loi de croissance logistique en régime discret.

Le comportement de la solution peut se déduire aisément de l'examen du graphique de N_{t+1} en fonction de N_k. À partir d'une valeur quelconque, de N_0, ce graphique permet de lire, en ordonnée, la valeur de N_1. Reportant cette valeur en abscisse, on en déduit la valeur N_2 suivante, ... En procédant de proche en proche, on construit une ligne brisée joignant alternativement des points de la courbe f à la bissectrice principale. Les abscisses des segments verticaux successifs correspondent à la suites des itérés $N_0, N_1, N_2, \ldots$

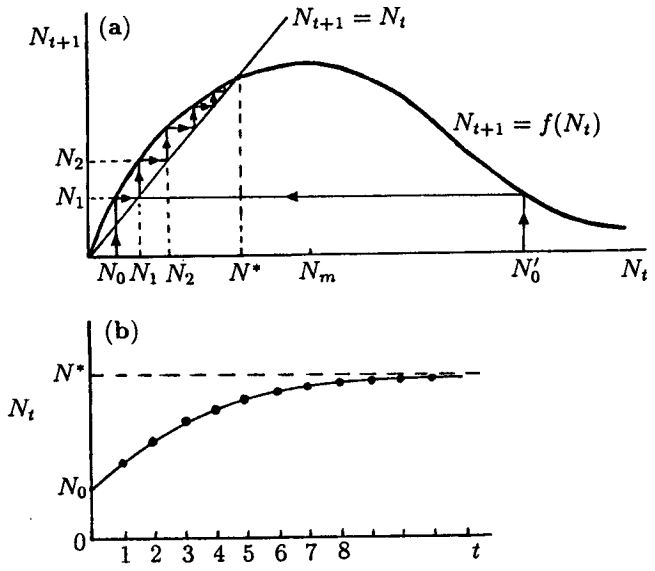


FIG. 5.7- Étude graphique des itérés d'une relation de récurence d'ordre un.

Si on excepte la valeur initiale N_0, toutes les valeurs N_k successives sont inférieures à la valeur maximale $N_{\max}$ de f. La population N est donc bornée.

L'intersection entre le graphique de f et la bissectrice principale en $N = N^*$ correspond à un état d'équilibre du système. En effet, on a alors

$$N_{k+1} = f(N_*) = N_*$$

et la population N_k est invariante au cours du temps.

Dans le cas illustré à la figure 5.7, l'équilibre en N^* est stable. Quelle que soit la perturbation de l'équilibre, $N_k \to N^*$ lorsque $t \to +\infty$. La convergence est monotone, sauf éventuellement lors du premier pas de temps si $N_0 > N^*$. Pour qualifier le fait que $N_k \to N^*$ quelle que soit la condition initiale N_0, on dit que l'espace entier constitue le bassin d'attraction de l'équilibre N^*.

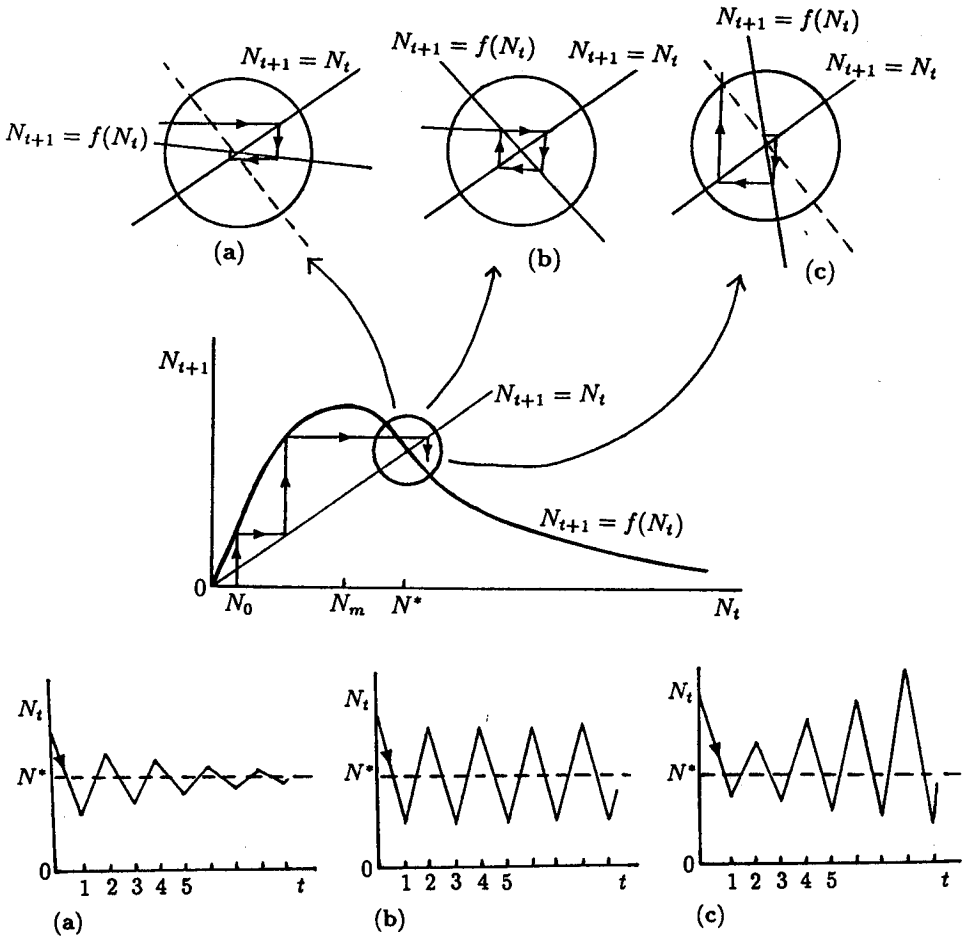


FIG. 5.8- Configurations stables et instables avec oscillation.

Lorsque les paramètres r et K donnent lieu à une représentation graphique comme à la figure 5.8-a, la convergence vers la situation d'équilibre s'accompagne d'oscillations. Dans le cas de la figure 5.8-c, l'équilibre est instable puisque le système s'écarte de la configuration d'équilibre (en présentant des oscillations).

Une étude analytique du comportement du système au voisinage de la position d'équilibre est tout à fait possible comme dans le cas des équations différentielles. Ainsi,

si

$$N_{t+1} = f(N_t) \tag{5.61}$$

les solutions d'équilibre sont les solutions de

$$N_* = f(N_*) \tag{5.62}$$

Introduisant la perturbation ε_k telle que

$$N_t = \varepsilon_t + N_* \tag{5.63}$$

et linéarisant (5.61) au voisinage de N_*, on trouve

$$\varepsilon_{t+1} + N_* = f(N_* + \varepsilon_t) \simeq f(N_*) + f'(N_*)\varepsilon_t \tag{5.64}$$

de sorte que l'évolution de la perturbation est décrite par

$$\varepsilon_{t+1} = f'(N_*)\varepsilon_t = [f'(N_*)]^{t+1}\varepsilon_0 \tag{5.65}$$

La stabilité de l'équilibre est donc déterminée par la valeur du paramètre $\lambda = f'(N_*)$.

- Si $|\lambda| < 1$, (avec $\lambda \neq 0$) la perturbation est amortie et l'équilibre est stable (pour autant que l'amplitude de la perturbation initiale ε_0 justifie la linéarisation). Le retour vers la position d'équilibre s'accompagne d'oscillations si $\lambda < 0$.
- Pour $|\lambda| > 1$, l'équilibre est instable (avec des oscillations d'amplitude croissante si $\lambda < -1$).
- Les cas $\lambda = \pm 1$ correspondent à des changements de comportement du système. On dit que celui-ci présente une bifurcation. Lorsque $\lambda = 1$, le graphique de f est tangent à la bissectrice principale en N^* de sorte que la bifurcation est dite tangente. Lorsque $\lambda = -1$, on parle de "bifurcation fourchette" ou "bifurcation par doublement de période".

Dans le cas du modèle de Ricker, les positions d'équilibre sont données par

$$N_* = N_* \exp\left[r\left(1 - \frac{N_*}{K}\right)\right] \qquad \Rightarrow \qquad \mathbb{N}_* \in \{0, K\} \tag{5.66}$$

L'origine est toujours instable puisque

$$f'(0) = e^r > 1 \quad \forall r > 0 \tag{5.67}$$

Pour déterminer la stabilité de $N_* = K$ (indépendant de r) on calcule

$$f'(N_*) = 1 - r \tag{5.68}$$

L'équilibre est donc stable pour $r \in]0, 2[$ et instable pour $r > 2$. La solution est oscillatoire pour $r > 1$.

L'instabilité de la solution pour $r > 2$ donne lieu à un comportement complexe. En effet, N_t ne peut tendre vers N_* puisque l'équilibre est instable, mais N_t ne peut plus croître au-delà de la valeur maximale de $f(N)$, soit

$$N_{\max} = \frac{K\, e^{r-1}}{r}$$

La solution oscille alors indéfiniment, et apparemment de façon aléatoire, autour de N_* (Fig. 5.9). Elle est dite *chaotique.*

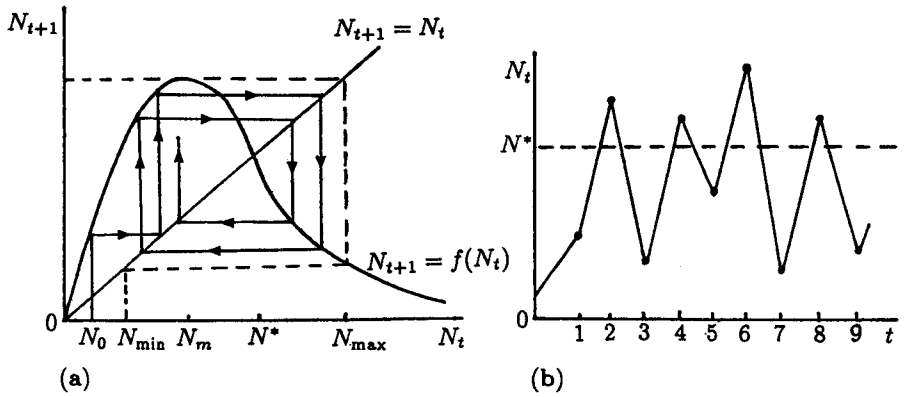


FIG. 5.9 – Comportement chaotique de l'équation de Ricker pour $r > 2$.

Sans entrer dans les détails, précisons que le régime chaotique n'est pas le domaine du hasard. Si le comportement d'un système chaotique semble erratique, c'est parce que son comportement ne se répète jamais et dépend très fortement des conditions initiales : des différences extrêmement faibles dans les valeurs des paramètres peuvent donner lieu à des résultats largement divergents. Un système chaotique n'en reste pas moins ordonné et déterministe : des effets objectifs et précisément mesurables et repérables déterminent univoquement la suite des événements. Le déterminisme est cependant qualifié d'imprévisible puisque, malgré la connaissance que nous avons de toutes les données qui déterminent les événements, l'extrême sensibilité aux conditions initiales nous empêche de dire ce qui va se passer.

Lorsque r est égal à la valeur limite $r_2 = 2$, le système ne possède plus aucun point d'équilibre stable. Le point $N_* = K$ est marginalement stable puisque les perturbations linéarisées vérifient

$$\varepsilon_{t+1} = -\varepsilon_t \tag{5.69}$$

La solution consiste donc en des oscillations de période 2. Pour des valeurs de r légèrement supérieures à $r_2 = 2$, de telles solutions périodiques de période 2 continuent

d'exister. Celles-ci correspondent à des points fixes de l'itération double

$$N_{t+2} = f[f(N_t)] \tag{5.70}$$

et peuvent donc être étudiées à partir de la solution (numérique) de

$$N = f[f(N)] \tag{5.71}$$

Dans le cas du modèle de Ricker, pour $r = 2.1$, on trouve par exemple une oscillation entre les valeurs $N_{2t} = 1.37$ et $N_{2t+1} = 0.63$. Ces points fixes de l'itération double sont stables, ce qui signifie que la solution périodique correspondant au passage de l'un à l'autre de ces points est elle-même stable.

Pour des valeurs de r supérieures à r_4, la solution de période 2 devient instable. Une solution périodique de période 4 apparaît. Celle-ci est stable dans une certaine gamme de valeurs de r et devient elle-même instable pour des valeurs de r supérieures à r_8.

Lorsque r augmente, le système passe au travers d'une série de bifurcations par doublement de période ; la solution stable de période p devenant instable alors qu'apparaît une solution stable de période $2p$. La distance entre les bifurcations dans l'espace-r devient de plus en plus petite. Au-delà d'une certaine valeur critique r_c, toutes les solutions périodiques de période 2^n sont instables. Le comportement du système devient alors extrêmement complexe : le chaos s'installe.

5.3.4 Modèle discret avec retard.

Tous les modèles discrets incorporent, par nature, la notion de retard. En effet, ils décrivent la taille de la population au temps t comme une fonction de la taille (au moins) à l'instant $t-1$ précédent. Comme dans le cas continu, le retard introduit un effet déstabilisateur dans les équations. Cet effet est d'autant plus grand que le retard est important. C'est pourquoi, même des modèles discrets apparemment très simples possèdent une dynamique complexe. Il n'est pas rare d'observer des oscillations de grande amplitude amenant les populations à des niveaux très faibles proches de l'extinction.

À titre d'exemple, considérons le modèle discret utilisé par la Commission Baleinière Internationale pour gérer la population des baleines. Désignant par N_t le nombre de baleines sexuellement matures au début de l'année t, on a

$$N_{t+1} = (1-\mu)N_t + R(N_{t-T}) \tag{5.72}$$

où le premier terme de membre de droite représente la population des baleines qui survivent d'une année à l'autre $(0 < \mu < 1)$ et le second terme modélise l'augmentation de la population adulte par les naissances intervenues T années plus tôt. Le délai T est celui de la maturité sexuelle et est de l'ordre de 5 à 10 ans. Le modèle suppose un sexe-ratio unitaire et une mortalité identique des deux sexes. Le terme de recrutement est de la forme

$$R(N) = \frac{1}{2}(1-\mu)^T N \left\{ P + Q\left[1 - \left(\frac{N}{K}\right)^z\right]\right\} \tag{5.73}$$

La constante K représente la population d'équilibre en dehors de toute chasse, P est la fécondité des femelles pour $N = K$, Q désigne l'augmentation maximale de la fécondité par laquelle l'espèce réagit lorsque la population est faible et z est un paramètre mesurant l'influence de cet effet. Le facteur $(1-\mu)^T$ tient compte du fait que les nouveaux-nés doivent survivre pendant T années avant leur maturité. Enfin le facteur $1/2$ tient compte du fait que la moitié seulement des baleines sont des femelles.

La constante K devant représenter la population d'équilibre, on doit avoir, posant $N_{t+1} = N_t = N_{t-T} = K$ dans (5.72),

$$\mu = \frac{1}{2}(1-\mu)^T P = h \tag{5.74}$$

Posant $q = \dfrac{Q}{P}$ et étudiant les perturbations de l'équilibre de la forme

$$N_t = K(1+\varepsilon_t) \tag{5.75}$$

il vient, après linéarisation,

$$\varepsilon_{t+1} = (1-\mu)\varepsilon_t + h(1-qZ)\varepsilon_{t-T} \tag{5.76}$$

Le polynôme caractéristique associé est de la forme

$$\lambda^{T+1} - (1-\mu)\lambda^T + h(1-q-z) = 0 \tag{5.77}$$

L'équilibre devient instable dès qu'un des zéros du polynôme vérifie $|\lambda| > 1$. Ces conditions peuvent être discutées en fonction de μ, T, h et qz. L'analyse, longue et compliquée, montre une forte sensibilité au paramètre z.

5.3.5 Modèle discret pour la gestion de la pêche.

Des modèles discrets de gestion de la pêche peuvent être construits et analysés comme dans le cas continu. Si la dynamique de la population est décrite par

$$N_{t+1} = f(N_t) \tag{5.78}$$

en l'absence de pêche et, si un prélèvement h_t est réalisé au temps t, alors

$$N_{t+1} = f(N_t) - h_t \tag{5.79}$$

À l'équilibre $N_t = N_{t+1} = N^*$ et

$$h^* = f(N^*) - N^* \tag{5.80}$$

En ce basant sur cet équilibre, le prélèvement durable maximum est obtenu pour $N^* = N_{max}$ solution de

$$\frac{dh^*}{dN^*} = f'(N_{max}) - 1 = 0$$
$$f(N_{max}) - N_{max} = h^*_{max}$$

La stratégie de gestion pourrait simplement être de maintenir la population N au niveau N_{max} correspondant au prélèvement maximum h^*_{max}. Cependant, le gestionnaire ne connaît en général pas la population réelle mais seulement l'amplitude des prises et une certaine mesure de l'effort de pêche (nombre d'heures en mer). Il importe donc de formuler le problème en terme d'effort de pêche et de prise.

En première approximation, on peut supposer que la prise par unité d'effort de pêche (par heure passée en mer) est proportionnelle à la population. Soit cN cette prise par unité d'effort. L'effort E_{max} nécessaire pour prélever $f(N_{max}) - N_{max} = h^*_{max}$ est donc donné par

$$E_{max} = \frac{1}{c} \int_{N_{max}}^{f(N_{max})} \frac{1}{N}\, dN = \frac{1}{c} \ln \frac{f(N_{max})}{N_{max}} \tag{5.81}$$

Cette équation constitue une relation paramétrique reliant l'effort au prélèvement en fonction de N_{max}. Elle peut être utile au gestionnaire pour limiter évaluer la population réelle en fonction du rapport prélèvement/effort et édicter des quotas de pêche.

Remarquons cependant que la discussion ci-dessus se base exclusivement sur l'équilibre. Si, à un moment donné, une augmentation de l'effort de pêche (ou, ce qui s'est produit dans plusieurs zones de pêches, une amélioration des techniques de pêche induisant une plus grande efficacité c) conduit à une réduction des prises, c'est que la prise durable maximale a été dépassée. Après réduction de l'effort de pêche, un certain temps peut être nécessaire pour que la population retrouve un niveau proche de N_{max}.

Notons encore que la gestion des ressources doit idéalement intégrer un aspect économique intégrant les coûts (en fonction de l'effort) et les bénéfices (en fonction de l'importance de la prise et des prix du marché).

Annexe - Complément sur les équations différentielles ordinaires.

Soit le problème différentiel

$$\begin{cases} \dot{y}_i(t) & = f_i(t, y_1, y_2, \ldots, y_n) \\ y_i(t_0) & = y_{0,i} \end{cases}, \qquad i = 1, 2, \ldots, n \tag{5.82}$$

à résoudre dans le domaine à $n+1$ dimensions

$$D = [t_0, t_0 + h] \times [y_{0,1} - \Delta_1, y_{0,1} + \Delta_1] \times \cdots \times [y_{0,n} - \Delta_n, y_{0,n} + \Delta_n] \tag{5.83}$$

Le théorème suivant donne des conditions suffisantes d'existence et d'unicité de la solution du problème (5.82) :

Si les fonctions f_i $(i = 1, 2, \ldots, n)$ sont continues sur D et telles que

- $|f_i(t, y_1, y_2, \ldots, y_n)| < M$ sur D ;
- il existe des constantes K_i $(i = 1, 2, \ldots, n)$ finies telles que

$$|\mathsf{f}(t, y_1, y_2, \ldots, y_n) - \mathsf{f}(t, y'_1, y'_2, \ldots, y'_n)| < K_1 |y_1 - y'_1| + \cdots + K_n |y_n - y'_n| \tag{5.84}$$

  pour tous les $(t, y_1, y_2, \ldots, y_n)$ et $(t, y'_1, y'_2, \ldots, y'_n)$ dans D (Condition de Lipschitz)

alors, le problème différentiel (5.82) possède une solution continue $(y_1(t), y_2(t))$ unique sur $t_0 \leq t \leq t_0 + h$ pour $h < \Delta_i / M$ $(i = 1, 2, \ldots, n)$.

Dans le cas d'un problème différentiel linéaire impliquant une seule fonction inconnue, on peut dégager des conditions suffisantes plus simples :

Si $a_{n-1}(x), a_{n-2}(x), \ldots, a_1(x), a_0(x)$ et $f(x)$ sont des fonctions continues sur l'intervalle `I`, alors il existe une fonction $y(x)$ unique n-fois continûment dérivable sur `I` qui satisfait à l'équation différentielle linéaire

$$\mathcal{L}_n(D)y(x) = y^{(n)}(x) + a_{n-1}(x)y^{(n-1)}(x) + \cdots + a_1(x)y'(x) + a_0(x)y(x) = f(x) \tag{5.85}$$

et qui vérifie les conditions initiales

$$y(x_0) = C_0, \quad y'(x_0) = C_1, \ldots, y^{(n-1)}(x_0) = C_{n-1} \tag{5.86}$$

où x_0 est un point arbitraire fixé dans `I` et où $C_0, C_1, \ldots, C_{n-1}$ sont des constantes quelconques.
La solution unique obtenu dépend continûment de x_0, C_0, C_1, …, C_{n-1}.

Selon ce dernier énoncé, la forme générale de la solution de (5.85) s'exprime au moyen de n constantes d'intégrations. Cette solution est appelée la *solution générale* de l'équation (5.85). Les n constantes d'intégration peuvent être fixées de façon unique par l'imposition de n conditions auxiliaires du type (5.86). On parle alors d'une *problème aux conditions initiales.* Les constantes d'intégration peuvent également (dans certains cas) être fixées par la données de n conditions auxiliaires imposées aux extrémités x_1 et x_2 de l'intervalle $[x_1, x_2]$ sur lequel le problème doit être résolu. Le problème est alors qualifié de *problème aux limites.*

Solution du problème linéaire.

Dans le cas d'un problème linéaire du type

$$y^{(n)}(x) + a_{n-1}(x)y^{(n-1)}(x) + \cdots + a_1(x)y'(x) + a_0(x)y(x) = f(x) \tag{5.87}$$

on montre que la solution générale peut s'écrire sous la forme

$$y(x) = y_h(x) + y_p(x) \tag{5.88}$$

où $y_h(x)$ désigne la solution générale de l'équation différentielle homogène correspondante

$$y^{(n)}(x) + a_{n-1}(x)y^{(n-1)}(x) + \cdots + a_1(x)y'(x) + a_0(x)y(x) = 0 \tag{5.89}$$

et $y_p(x)$ représente une solution particulière quelconque de l'équation non-homogène de départ.

En pratique, la recherche de la solution d'un problème différentiel linéaire se décompose donc en quatre étapes :

– recherche de la solution générale de l'équation homogène,
– recherche d'une solution particulière de l'équation non homogène.
– formation de la solution générale de l'équation non homogène en ajoutant la solution particulière de l'équation non homogène et la solution générale de l'équation homogène,
– détermination des constantes d'intégration apparaissant dans la solution générale en exploitant les conditions initiales ou aux limites.

Solution générale des équations linéaires à coefficients constants.

Dans le cas particulier de l'équation différentielle linéaire

$$a_n y^{(n)}(x) + a_{n-1} y^{(n-1)}(x) + \cdots + a_1 y'(x) + a_0 y(x) = f(x) \tag{5.90}$$

où les coefficients $a_n, a_{n-1}, \ldots, a_1, a_0$ sont constants ($a_n \neq 0$) et où la fonction $f(x)$ est continue sur un intervalle `I`, on dispose de méthodes systématiques de résolution.

Pour déterminer la solution générale de l'équation homogène

$$a_n y^{(n)}(x) + a_{n-1} y^{(n-1)}(x) + \cdots + a_1 y'(x) + a_0 y(x) = 0 \tag{5.91}$$

on commence par construire le *polynôme caractéristique* associé à cette équation, soit

$$\mathcal{L}_n(\lambda) = a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 \tag{5.92}$$

La solution générale de l'équation homogène est alors obtenue à partir de l'étude des zéros du polynôme caractéristique :

Si le polynôme caractéristique $\mathcal{L}_n(z)$ possède m zéros distincts λ_i de multiplicité α_i avec $i = 1, \ldots, m$ et $\sum_{i=1}^{m} \alpha_i = n$, alors la solution générale de l'équation homogène (5.91) s'écrit

$$y_h(x) = \sum_{i=1}^{m} \mathcal{P}_{\alpha_i - 1}(x)\, \mathrm{e}^{\lambda_i x} \tag{5.93}$$

où $\mathcal{P}_{\alpha_i - 1}(x)$ est un polynôme de degré $\alpha_i - 1$.

Des méthodes systématiques existent également pour déterminer une solution particulière du problème non-homogène (5.90). Dans les cas simples, on peut généralement déterminer une telle solution – ou du moins sa forme paramétrique – par simple inspection. Ainsi, dans le cas particulier ou le second membre s'écrit sous la forme

$$f(x) = \mathcal{P}_p(x) e^{\beta x} \tag{5.94}$$

où $\mathcal{P}_p(x)$ un polynôme de degré p, on montre que l'équation (5.90) admet une solution particulière de la forme

$$y_p(x) = x^{\alpha}\mathcal{P}_p^*(x)\, e^{\beta x} \tag{5.95}$$

où α désigne la multiplicité de β comme zéro du polynôme caractéristique $\mathcal{L}_n(\lambda)$ et $\mathcal{P}_p^*$ désigne un polynôme de degré p (en général différent de $\mathcal{P}_p$). Il ne reste plus alors qu'à identifier les coefficients du polynôme inconnu $\mathcal{P}_p^*$ en substituant (5.95) dans l'équation (5.90).

EXEMPLE 5.1 Considérons le mouvement d'une corps de masse volumique $\rho^\star$ et de volume V dans une colonne d'eau stratifiée. Soit z la coordonnée verticale (positive vers le haut) et $\rho(z)$ la distribution verticale de la masse volumique de l'eau. L'équation du mouvement vertical du corps s'obtient par application de la loi de Newton en tenant compte du poids du corps et de la poussée d'Archimède, soit

$$\rho^\star V\frac{d^2z}{dt^2} = -\rho^\star Vg + \rho(z)Vg$$

soit

$$\frac{d^2z}{dt^2} = \left(\frac{\rho(z)}{\rho^\star} - 1\right) g$$

Linéarisant la fonction $\rho(z)$ au voisinage du point $z^\star$ pour lequel $\rho(z^\star) = \rho^\star$, on a

$$\rho(z) \approx \rho^\star + \left.\frac{d\rho}{dz}\right|_{z=z^\star} (z - z^\star)$$

et

$$\frac{d^2z}{dt^2} = \frac{g}{\rho^\star}\left.\frac{d\rho}{dz}\right|_{z=z^\star} (z - z^\star)$$

Introduisant la fréquence de Brunt-Väisälä N telle que

$$N^2 = -\frac{g}{\rho^\star}\left.\frac{d\rho}{dz}\right|_{z=z^\star}$$

(la densité diminue lorsque z augmente si la colonne d'eau est stable), l'équation devient

$$\frac{d^2z}{dt^2} + N^2z = N^2z^\star$$

La solution générale de l'équation homogène

$$\frac{d^2z}{dt^2} + N^2z = 0$$

est obtenue en identifiant les zéros du polynôme caractéristique

$$\lambda^2 + N^2 = 0$$

soit

$$\lambda_1 = -iN, \qquad \lambda_2 = iN$$

On a donc

$$z_h(t) = C_1 \exp -iNt + C_2 \exp iNt$$

où C_1 et C_2 sont des constantes d'intégration quelconques. Utilisant la correspondance entre les fonctions trigonométriques et les exponentielles imaginaires,

$$e^{ix} = \cos x + i \sin x, \qquad \cos x = \frac{e^x + e^{-x}}{2}, \qquad \sin x = \frac{e^x + e^{-x}}{2i}$$

, on peut écrire la solution sous la forme

$$z_h(t) = \tilde{C}_1 \sin(Nt) + \tilde{C}_2 \cos(Nt)$$

où $\tilde{C}_1$ et $\tilde{C}_2$ sont de nouvelles constantes inconnues.

Selon (5.95), une solution particulière de l'équation non-homogène

$$\frac{d^2z}{dt^2} + N^2 z = N^2 z^\star$$

peut être exprimée sous la forme (prenant $\beta = 0 = \alpha$)

$$z_p(t) = \mathcal{P}_0^*(t) = C$$

où C désigne une constante. Substituant cette expression paramétrique dans l'équation, on trouve aisément

$$C = z^\star.$$

La solution générale du problème est donc

$$z(t) = z_h(t) + z_p(t) = \tilde{C}_1 \sin(Nt) + \tilde{C}_2 \cos(Nt) + z^\star$$

Si le corps est lâché sans vitesse à une hauteur h au-dessus de sa position d'équilibre $z^\star$, il vient

$$\begin{cases} z(0) = z^\star + h = \tilde{C}_2 + z^\star \\ z'(0) = 0 = \tilde{C}_1 N \end{cases}$$

Ces conditions permettent de fixer la valeur des constantes d'intégration

$$\begin{cases} \tilde{C}_1 = 0 \\ \tilde{C}_2 = h \end{cases}$$

et la solution complète du problème est donc donnée par

$$z(t) = z^\star + h \cos(Nt)$$

Le corps oscille donc autour de sa position d'équilibre $z^\star$ avec une pulsation N. ◇

Chapitre 6

Modélisation dynamique avec interactions.

Les problèmes environnementaux sont généralement caractérisés par des interactions fortes entre leurs différentes composantes, qu'il s'agisse d'espèces chimiques qui réagissent lorsqu'elles sont mises en contact l'une avec l'autre ou d'espèces différentes d'un réseau trophique qui se nourrisent l'une de l'autre. Par une modélisation mathématique adaptée, on peut décrire ces interactions pour en décrire la dynamique, les configurations d'équilibre et les instabilités. Les modèles mathématiques les plus simples comprennent un système d'équations différentielles qui sont couplées par des termes souvent non linéaires.

Cette section présente les concepts et méthodes mathématiques d'analyse fondamentaux applicables à ce type de modèle.

6.1 Modèles continus.

6.1.1 Modélisation des transformations biochimiques.

Selon la loi d'action des masses de Guldberg et Waage, la vitesse d'une réaction chimique est proportionnelle au produit des concentrations des réactifs élevées à une puissance égale au coefficient stoechiométrique correspondant. Pour une réaction du type

$$\alpha A + \beta B \rightarrow \gamma C + \delta D \tag{6.1}$$

où $\alpha, \beta, \gamma, \delta$ sont les coefficients stoechiométriques des réactifs A et B et des produits C et D, la vitesse de réaction est donnée par

$$v^+ = k^+ [A]^\alpha [\beta]^\beta \tag{6.2}$$

où k^+ désigne la constante de vitesse de la réaction. Lorsque la réaction progresse d'une unité vers la droite, γ môles de C sont produites de sorte que

$$\frac{d[C]}{dt} = \gamma v^+ = \gamma k^+ [A]^\alpha [\beta]^\beta \tag{6.3}$$

De même

$$\frac{d[D]}{dt} = \delta v^+, \frac{d[A]}{dt} = -\alpha V^+; \frac{d[B]}{dt} = -\beta v^+ \tag{6.4}$$

Si la réaction inverse est également possible, le système tend vers l'équilibre

$$\alpha A + \beta B \rightleftharpoons \gamma C + \delta D \tag{6.5}$$

La vitesse de la réaction de droite à gauche est donnée par

$$v^- = k^-[C]^\gamma[D]^\delta \tag{6.6}$$

où k^- désigne la constante de vitesse correspondante. Les concentrations des différents réactifs et produits évoluent selon

$$\frac{d[A]}{dt} = \alpha(v^- - v^+) \tag{6.7}$$

$$\frac{d[B]}{dt} = \beta(v^- - v^+) \tag{6.8}$$

$$\frac{d[C]}{dt} = \gamma(v^+ - v^-) \tag{6.9}$$

$$\frac{d[D]}{dt} = \delta(v^+ - v^-) \tag{6.10}$$

L'équilibre est atteint lorsque

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = \frac{d[C]}{dt} = \frac{d[D]}{dt} = 0 \tag{6.11}$$

c'est-à-dire

$$v^+ = v^- \tag{6.12}$$

soit

$$\frac{[C]^\gamma[D]^\delta}{[A]^\alpha[B]^\beta} = \frac{k^+}{k^-} = K_{eq} \tag{6.13}$$

On retrouve donc la loi d'équilibre habituelle de constante d'équilibre K_{eq}.

Les constantes de vitesse k^+ et k^- des réactions ont des grandeurs (et des unités) qui dépendent des ordres de des réactions. Pour une relation du type

$$\frac{d[A]}{dt} = -k[A]$$

on a

$$[k] = T^{-1}$$

Pour une réaction d'ordre P, on a

$$[k] = (M\,L^{-3})^{1-p}T^{-1}$$

Les constantes de vitesse dépendent de la température selon une loi du type

$$k_{T_1} = k_{T_2} \exp\left[\frac{E_{\text{act}}}{RT_1T_2}(T_1 - T_2)\right]$$

où k_{T_1} et k_{T_2} désignent les constantes de réaction aux températures absolues T_1 et T_2, E_{act} est l'énergie d'activation de la réaction et R la constante universelle des gaz parfaits (8.314 J mol$^{-1}K^{-1}$).

Dans le domaine de température relativement restreint $(0^\circ - 35^\circ)$ dans lequel se placent les études environnementales, on remplace la relation précédente par

$$k_{T^\circ} = k_{20}\theta^{T^\circ - 20} \tag{6.14}$$

où θ est une constante supérieure à l'unité (de l'ordre de 1.0 à 1.1 pour beaucoup de réactions) où T° désigne la température en $^\circ C$ et k_{20} est la constante de réaction à 20°C. Une forme semblable à (6.14) est également souvent utilisée pour représenter la dépendance du taux d'activité des bactéries ou du zoo-plancton en fonction de la température.

Solutions de base pour des réactions simples d'ordre 0, 1 et 2.

Réaction d'ordre 1

Une réaction d'ordre un est une réaction du type

$$A \rightarrow B \tag{6.15}$$

Ce type de réaction décrit, par exemple, les processus de décroissance radioactive ou de mortalité et de respiration des bactéries et des algues.

On a

$$\frac{d[A]}{dt} = -k[A] = -\frac{d[B]}{dt}$$

Par intégration à partir de conditions initiales $[A]_0$ et $[B]_0$, il vient

$$\begin{aligned} [A](t) &= [A]_0\, e^{-kt} \\ [B](t) &= [B]_0 + [A]_0\,(1 - e^{-kt}) \end{aligned} \tag{6.16}$$

Lorsque le mécanisme d'une réaction est inconnu, une réaction du premier ordre (6.15) peut être proposé si les concentrations des réactifs et des produits varient exponentiellement comme dans (6.16).

De façon équivalente, on reconnaît une réaction d'ordre un peu le fait que les courbes des concentrations en fonction du temps apparaissent comme des droites sur une échelle logarithmique.

Réactions d'ordre 2

La plupart des réactions du second ordre de la chimie aquatique sont d'un des types

$$A+A \rightarrow B \tag{6.17}$$
$$A+B \rightarrow C \tag{6.18}$$
$$A+R \rightarrow 2R \tag{6.19}$$

La dernière équation correspond à une relation auto-catalytique.

L'équation (6.17) donne lieu à

$$\begin{cases} \dfrac{d[A]}{dt} = -2k[A]^2 \\ \dfrac{d[B]}{dt} = k[A]^2 \end{cases}$$

donc

$$[A](t) = \frac{[A]_0}{1+2kt[A]_0}, \quad [B](t) = [B]_0 + \frac{kt[A]_0^2}{1+2kt[A]_0}$$

L'évolution de la concentration du réactif A est telle que $1/[A](t)$ est une fonction linéaire croissante de t.

Dans le cas de (6.18), on a

$$\frac{d[A]}{dt} = -k[A][B]$$

De même

$$\frac{d[B]}{dt} = -k[A][B]$$

En prenant la différence de ces deux équations, on a

$$\frac{d}{dt}([A]-[B]) = 0$$

et

$$[B](t) = [A](t) + [B]_0 - [A]_0$$

Dès lors

$$\frac{d[A]}{dt} = -k[A]([A]+[B]_0-[A]_0)$$

et

$$\int_{[A]_0}^{[A](t)} \frac{d[A']}{[A']([A']+[B]_0-[A_0]_0)} = \int_0^t -k\,dt'$$

Soit

$$[A](t) = \frac{[A]_0([A]_0-[B]_0)}{[A]_0-[B]_0\exp[-kt([A]_0-[B]_0)]}$$
$$[B](t) = \frac{[B]_0([B]_0-[A]_0)}{[B]_0-[A]_0\exp[-kt([B]_0-[A]_0)]}$$

tant que $[A](t), [B](t) \geq 0$.

Une réaction de ce type peut être identifiée expérimentalement en remarquant que

$$\ln \frac{[A](t)}{[B](t)} = \ln \frac{[A]_0}{[B]_0} - k([B]_0 - [A_0])t$$

i.e. que $\ln[A]/[B]$ évolue linéairement au cours du temps. Une représentation de cette grandeur en fonction de t permet donc d'identifier la pente de la droite avec $-k([B]_0 - [A]_0)$.

La réaction auto-catalytique est décrite par les équations différentielles

$$\begin{aligned} \frac{d[A]}{dt} &= -k[A][R] \\ \frac{d[R]}{dt} &= k[A][R] \end{aligned}$$

et peut donc être traitée comme (6.18).

Les réactions d'ordre 3 ou plus sont rares. Elles correspondraient en effet à des interactions de 3 molécules. Lorsqu'on écrit une réaction de ce type, il s'agit généralement de l'écriture compacte d'une suite de réactions simples. Si la loi d'action des masses est valable pour chacune des réactions simples, elle n'est cependant pas valable pour la réaction totale sans approximation. Il arrive même souvent que la réaction totale puisse être approchée par une loi d'ordre zéro du type

$$\frac{d[A]}{dt} = -k_0.$$

C'est le cas, par exemple, de la production de méthane et la libération des produits d'hydrolyse (NH_3, PO_3^-) dans une couche anaérobique des sédiments.

6.1.2 Réactions composées

En raison de la grande diversité des espèces chimiques présentes dans l'environnement, les transformations chimiques sont généralement réalisées par le biais d'une suite de réactions chimiques sucessives dont un grand nombre sont irréversibles.

Considérons tout d'abord une séquence de réactions du premier ordre

$$A \xrightarrow{k} B \xrightarrow{k_2} C \tag{6.20}$$

typique de la désintégration radioactive d'un élément A en éléments de plus en plus légers

B et C. On a

$$\frac{d[A]}{dt} = -k_1[A] \tag{6.21}$$

$$\frac{d[B]}{dt} = -k_1[A] - k_2[B] \tag{6.22}$$

$$\frac{d[C]}{dt} = -k_2[B] \tag{6.23}$$

Remarquons que

$$\frac{d}{dt}([A]+[B]+[C]) = 0$$

de sorte que

$$[A]+[B]+[C] = \text{Constante}$$

Les équations (6.21) - (6.23) peuvent être résolues successivement. Si on suppose qu'initialement $[B]_0 - [C]_0 = 0$, alors

$$\begin{aligned}
[A] &= [A]_0 e^{-k_1 t} \\
[B] &= \frac{k_1}{k_1 - k_2}[A]_0 (e^{-k_2 t} - e_{k_1 t}) \\
[C] &= [A]_0 \left(1 + \frac{k_2\, e^{-k_1 t} - k_1\, e_{k_2 t}}{k_1 - k_2}\right)
\end{aligned}$$

Si $k_1 \gg k_2$, ces expressions peuvent être approchées par

$$\begin{aligned}
[A] &= [A]_0 e^{-k_1 t} \\
[B] &\simeq [A]_0 e^{-k_2 t} \\
[C] &\simeq [A]_0 (1 - e^{-k_2 t})
\end{aligned}$$

La séquence des réactions (6.20) peut donc elle-même être décrite par

$$A \xrightarrow{k_2} C$$

Inversement, si $k_2 \gg k_1$, on a

$$\begin{aligned}
[A] &= [A]_0 \cdot e^{-k_1 t} \\
[B] &\simeq \frac{k_1}{k_2}[A]_0 e^{-k_1 t} \\
[C] &\simeq [A]_0 (1 - e^{-k_1 t})
\end{aligned}$$

ce qui correspond à

$$A \xrightarrow{k_1} C$$

Dans les deux cas, on constate que la dynamique de la réaction globale est dictée par la cinétique de la réaction la plus lente.

Il s'agit là d'un principe tout à fait général de modélisation des systèmes dynamiques. Il faut être particulièrement attentif à paramétrer précisément la dynamique des processus les plus lents car ce sont eux qui gouvernent la dynamique globale. Les processus les plus rapides peuvent par contre être considérés comme immédiats sans engendrer d'erreur importante. Les variables d'états liées par les processus très rapides peuvent elles-mêmes être groupées pour former un agrégat unique, diminuant ainsi la dimension du problème. Ainsi, dans le cas où $k_1 \gg k_2$, il est inutile de traiter séparément les espèces A et B. La transformation de A en B étant quasi-immédiate, on ne distinguera pas ces deux espèces.

Le schéma de réaction (6.20) s'applique particulièrement bien à l'oxydation de l'ammonium en nitrite puis nitrate par le biais des réactions successives

$$NH_3 + \frac{3}{2}O_2 \rightarrow NO_2^- + H^+ + H_2O \tag{6.24}$$

$$NO_2^- + \frac{1}{2}O_2 \rightarrow NO_3^- \tag{6.25}$$

Ces transformations sont équivalentes à la transformation globale

$$NH_3 + 2O_2 \rightarrow NO_3^- + H^+ H_2O \tag{6.26}$$

(Remarquons les bactéries Nitrosomas spp. et Nitrobacter spp. interviennent respectivement comme catalyseurs de (6.24) et (6.25).)

La concentration en oxygène n'affectant par la cinétique des réactions (sauf si les niveaux d'oxygène sont très faibles), la cinétique des réactions peut être décrite par (6.21) - (6.23).

6.1.3 Réactions réversibles

La plupart des réactions acide-base, de complexion ou d'adsorption-désorption, sont réversibles ; des modifications de concentrations ou de conditions environnementales peuvent alors engendrer un déplacement de l'équilibre. Le schéma de base pour une telle réaction est

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B$$

La cinétique de la réaction est décrite par

$$\frac{d[A]}{dt} = -k_1[A] + k_2[B] = -\frac{d[B]}{dt}$$

La concentration totale des deux espèces A et B est donc constante

$$[A] + [B] = C = [A]_0 + [B]_0$$

Dès lors

$$\frac{d[A]}{dt} + (k_1 + k_2)[A] = k_2 C$$

et

$$[A](t) = \frac{k_2 C}{k_1 + k_2}(1 - e^{-(k_1+k_2)t}) + [A]_0 e^{-(k_1+k_2)t}$$

De même

$$[B](t) = \frac{k_1 C}{k_1 + k_2}(1 - e^{-(k_1+k_2)t}) + [B]_0 e^{-(k_1+k_2)t}$$

Lorsque t augmente, $[A]$ $[B]$ tendent vers leurs valeurs d'équilibre

$$[A]_\infty = \frac{k_2 C}{k_1 + k_2}, \quad [B]_\infty = \frac{k_1 C}{k_1 + k_2}$$

qui sont bien telles que

$$\frac{[B]_\infty}{[A]_\infty} = \frac{k_1}{k_2} = K_{eq}$$

D'un point de vue dynamique, on constate que la convergence vers l'état d'équilibre est caractérisée par $(k_1 + k_2)$ ou, de façon équivalente, par le temps caractéristique

$$\tau = \frac{1}{k_1 + k_2}$$

Si on s'intéresse aux échelles de temps bien supérieures à τ, l'équilibre peut être supposé réalisé en bonne approximation. Par contre, pour $t = o(\tau)$, l'aspect dynamique ne peut être négligé.

6.1.4 Réaction enzymatique

Un grand nombre de réactions font intervenir des protéines spéciales, appelées enzymes, qui agissent comme des catalyseurs très efficaces des ces réactions. Les enzymes réagissent sélectivement à certains substrats et régulent les processus biologiques.

La réaction enzymatique de base, proposée pour la première fois par Michaelis et Menten (1913), comporte deux étapes selon le schéma

$$E + S \underset{k_1'}{\overset{k_1}{\rightleftharpoons}} SE \overset{k_2}{\to} P + E \tag{6.27}$$

L'enzyme E et le substrat S se combinent pour former un complexe SE qui est lui-même converti en un produit P. Lors de cette dernière étape, l'enzyme est également reformée de sorte que l'enzyme n'est pas consommée par la réaction mais permet seulement d'augmenter la vitesse de la transformation globale. L'enzyme E et le substrat S sont en équilibre avec leur complexe SE. La constante de réaction k_2 est généralement petite par rapport à k_1 et k_1'.

La loi d'action des masses appliquée à (6.27) permet d'écrire

$$\frac{d[SE]}{dt} = k_1[E]\,[S] - k_1'[SE] - k_2[SE] \tag{6.28}$$

$$\frac{d[S]}{dt} = -k_1[E]\,[S] + k_1'[SE] \tag{6.29}$$

$$\frac{d[E]}{dt} = -k_1[E]\,[S] + k_1'[SE] + k_2[SE] \tag{6.30}$$

$$\frac{d[P]}{dt} = k_2[SE] \tag{6.31}$$

Ces équations, complétées par des conditions initiales appropriées, permettent de déterminer complètement l'évolution des concentrations des différentes espèces.

L'approche classique consiste à supposer que

$$\frac{d[SE]}{dt} \simeq 0 \tag{6.32}$$

en s'appuyant sur la cinétique rapide de la première réaction et la réalisation quasi-immédiate de l'équilibre. Dans ce cas,

$$k_1[E]\,[S] \simeq (k_1' + k_2)[SE] \tag{6.33}$$

et, puisque k_2 est négligeable par rapport à k_1',

$$\frac{[E]\,[S]}{[SE]} = \frac{k_1'}{k_1} = K_{eq} \tag{6.34}$$

D'autre part, en combinant les équations (6.28) et (6.30), on vérifie aisément que

$$[E] + [SE] = e_0 \tag{6.35}$$

où e_0 est une constante : la somme des concentrations des formes libre et combinée de l'enzyme est constante.

En combinant les relations (6.34) et (6.35), on obtient

$$[SE] = \frac{e_0[S]}{K_{eq} + [S]} \tag{6.36}$$

Dès lors,

$$\frac{d[P]}{dt} = k_2 \frac{e_0[S]}{K_{eq} + [S]} \tag{6.37}$$

On constate que le taux de production de P est proportionnel à la concentration totale e_0 de l'enzyme. Celle-ci agit donc bien comme catalyseur.

Dans le cas où la réaction (6.28) est celle de la synthèse cellulaire, P désigne la biomasse produite et $k_2 e_0$ représente le taux de croissance maximum. Exprimant ce dernier sous la forme $k_2 e_0 = \mu_{\max}[P]$, on retrouve la loi classique de Michaelis-Menten

$$\frac{d[P]}{dt} = \mu_{\max} \frac{[P]\,[S]}{K_{eq} + [S]} \tag{6.38}$$

Cette expression n'est caractéristique ni d'une réaction de premier ordre, ni d'une réaction du second ordre. Elle est intermédiaire entre ces deux situations puisque, si $[S] \ll K_{eq}$, on a

$$\frac{d[P]}{dt} = \mu_{\max}[P]$$

qui est du premier ordre, et, si $[S] \gg K_{eq}$, la saturation se fait sentir,

$$\frac{d[P]}{dt} \simeq \frac{\mu_{\max}}{K_{eq}}[P]\,[S]$$

et la réaction est du second ordre.

La simulation numérique de l'évolution des concentrations des différentes espèces permet de mettre en évidence les conditions de validité de l'hypothèse (6.32). Si les concentrations initiales de P et SE sont nulles, on observe une première phase très courte pendant laquelle les concentrations du substrat $[S]$ et du produit $[P]$ sont très peu modifiées tandis que les concentrations de $[E]$ et $[SE]$ varient rapidement pour atteindre un pseudo-équilibre.

Dans une seconde phase, les concentrations des différentes espèces varient lentement jusqu'à la transformation complète du substrat en produit P. La durée de la première phase est de l'ordre de

$$t_c = \frac{1}{k_1(S_0 + K_M)} \tag{6.39}$$

où $S_0 = [S](0)$ désigne la concentration initiale du substrat. En effet, la phase initiale est caractérisée par le fait que $[S]$ reste pratiquement constant. Dès lors, (6.28) peut être approché par

$$\begin{aligned}
\frac{d[SE]}{dt} &\simeq k_1 S_0 [E] - k_1' [SE] \\
&\simeq k_1 S_0 [(e_0 - [SE]) - k_1' [SE] \\
&\simeq k_1 S_0 e_0 - k_1 (S_0 + K_{eq}) [SE]
\end{aligned}$$

Le temps caractéristique de la croissance exponentielle de $[SE]$ (et de la décroissance correspondante de $[E]$) est donc donné par (6.39). Dans beaucoup de situations expérimentales, ce temps est très court et le comportement rapide du système n'est pas observable. En tout cas, la première phase peut être ignorée et le système peut être raisonnablement décrit par (6.37) ou (6.38) si on s'intéresse uniquement à des échelles de temps bien supérieures à t_c.

6.1.5 Modèle proie-prédateur de Lotka-Volterra.

Le modèle de Lotka-Volterra est un modèle classique utilisé pour décrire l'interaction entre des populations de proies et de prédateurs, les premiers se nourrissant des seconds. Initialement présenté par Lotka (1920) pour décrire la dynamique d'une réaction chimique, ce modèle a été également proposé par Volterra (1926) pour décrire les oscillations temporelles de certaines espèces de poissons en mer Adriatique.

Notons $N(t)$ la population de la proie au temps t et $P(t)$ la population de son prédateur. Dans le modèle de base de Lotka-Volterra, on suppose d'une part que la population de la proie se développe exponentiellement en l'absence de prédation. Celle-ci est proportionnelle à la probabilité de rencontre entre les deux espèces et donc également proportionnelle au produits des deux populations. D'autre part, le prédateur est incapable de se maintenir sans proie car il connaît un taux de mortalité constant. Ces hypothèses peuvent être traduites mathématiquement par

$$\begin{cases} \dfrac{dN}{dt} = aN - bN\,P \\ \dfrac{dP}{dt} = cN\,P - dP \end{cases} \tag{6.40}$$

Assorti de conditions auxiliaires appropriées décrivant les populations initiales N_0 et P_0 de la proie et du prédateur, ce système permet, par intégration (numérique), de décrire le populations des deux espèces à tous les instants ultérieurs.

Afin d'étudier la dynamique de (6.40), il est avantageux d'introduire les variables adimensionnelles

$$u(\tau) = \frac{cN(t)}{d}, \qquad v(\tau) = \frac{bP(t)}{a}, \qquad \tau = at, \qquad \alpha = \frac{d}{a} \tag{6.41}$$

en fonction desquelles le système (6.40) prend la forme

$$\frac{du}{d\tau} = u(1-v), \qquad \frac{dv}{d\tau} = \alpha v(u-1) \tag{6.42}$$

Cette expression du modèle fait clairement apparaître que, aux variations de l'échelle temporelle et aux mises à échelles des variables décrivant les deux populations près, la dynamique du système dépend du seul paramètre adimensionnel α représentant le rapport du taux de mortalité du prédateur et du taux de croissance de la proie.

La résolution analytique complète du système (6.42) n'est pas possible. Cependant, on peut obtenir des informations importantes sur la dynamique du système en examinant l'existence et la nature des solutions particulières correspondant à des valeurs constantes de u et v. Celles-ci correspondent à l'annulation des seconds membres de (6.42) et sont donc données par

$$(u,v) = (0,0) \qquad \text{et} \qquad (u,v) = (1,1) \tag{6.43}$$

Ces couples définissent les *points critiques* ou *points d'équilibre* du système étudié. Si le système est abandonné dans une telle configuration à un moment donnée, alors les

dérivées temporelles de u et v sont nulles et le système demeure dans cet état indéfiniment. Le point d'équilibre $(1,1)$, en particulier, décrit les populations constantes de la proie et du prédateur qui peuvent cohabiter dans le système étudié.

Au delà de l'existence des points critiques, il importe d'en définir la nature en effectuant une analyse de stabilité locale. Pour ce faire, comme dans le cas unidimensionnel, on linéarise les équations au voisinage du point étudié.

Considérons tout d'abord le point critique $(0,0)$ correspondant à l'absence de proie et de prédateur. Considérant une petite perturbation de cet état, la linéarisation des équations (6.42) conduit à

$$\frac{du}{d\tau} \approx u, \qquad \frac{dv}{d\tau} \approx -\alpha v \tag{6.44}$$

dont la solution est donnée par

$$u(\tau) = C_1 \, \mathrm{e}^{\tau}, \qquad v(\tau) = C_2 \, \mathrm{e}^{-\alpha\tau} \tag{6.45}$$

Celle-ci correspond à l'extinction de la population de prédateur et la croissance exponentielle de la proie. Puisque l'une au moins de populations croit exponentiellement, le point d'équilibre $(0,0)$ est qualifié d'instable. Puisque toutes les variables ne présentent pas le même comportement, *i.e.* que $u(\tau)$ croit exponentiellement et $v(\tau)$ décroît exponentiellement, le point critique est appelé un *point de selle.*

Une analyse de stabilité semblable peut être réalisée pour le second point critique $(1,1)$. Introduisant cette fois explicitement les perturbations $\varepsilon(\tau)$ et $\eta(\tau)$ telles que

$$u(\tau) = 1 + \xi(\tau), \qquad v(\tau) = 1 + \eta(\tau) \tag{6.46}$$

et supposant que ces perturbations sont faibles devant l'unité, l'évolution des perturbations au voisinage du point critique peut être approchée par

$$\frac{d\xi}{d\tau} = -(1+\xi)\eta \approx -\eta, \qquad \frac{d\eta}{d\tau} = \alpha\xi(1+\eta) \approx \alpha\xi \tag{6.47}$$

En dérivant la première relation approchée et en remplaçant la dérivée de η par sa valeur approchée extraite de la seconde équation, on obtient l'équation du second ordre

$$\frac{d^2\xi}{d\tau} + \alpha\xi = 0 \tag{6.48}$$

qui décrit des oscillations harmoniques de la perturbation. La perturbation η vérifiant une équation semblable, on en déduit que le comportement du système au voisinage du point critique étudié est constitué d'oscillations harmoniques non amorties de période $2\pi/\sqrt{\alpha}$ autour de $(1,1)$. On dit alors que $(1,1)$ constitue un *centre* et que l'équilibre est *marginalement stable.* Puisque les perturbations restent bornées, l'équilibre est *stable.* Cependant, comme les perturbations ne sont pas amorties au cours du temps, la stabilité est dite *marginale*; de petites perturbations ou les termes non linéaires négligés dans l'analyse locale de stabilité pourraient engendrer une lente croissance ou décroissance de l'amplitude des perturbations.

Les équations linéarisées (6.44)-(6.47) permettent de décrire le comportement du système au voisinage des points critiques. Dans le cas d'une point d'équilibre stable, l'approximation induite par la linéarisation est consistante avec la solution obtenue ; si la perturbation initiale est suffisamment petite pour justifier la linéarisation, la linéarisation restera appropriée aux instants ultérieurs. Dans le cas d'un point d'équilibre instable, la croissance d'une des variables (au moins) met à mal la validité de la linéarisation. Les équations linéarisées ne permettent pas de rendre compte du comportement du système lorsque les perturbations deviennent grandes. Dans les deux cas, on ne peut rien dire au sujet du comportement du système en dehors des voisinages des points critiques.

Pour décrire globalement le comportement du système, on peut représenter l'état de celui-ci dans le plan-(u,v) ; chaque point du plan étant représentatif d'un état du système. Au cours du temps, (u,v) décrit une courbe orientée dans ce plan. Cette courbe est appelée la *trajectoire de phase* du système tandis que le plan (u,v) est appelé le *plan de phase.*

Les solutions approchées obtenues précédemment par linéarisation permettent de dessiner les trajectoires au voisinage des points critiques. Celles-ci prennent l'allure d'hyperboles au voisinage de l'origine et d'ellipses au voisinage du point d'équilibre stable $(1,1)$.

Les trajectoires de phase exactes sont décrites par l'équation différentielle

$$\frac{dv}{du} = \alpha \frac{v(u-1)}{u(1-v)} \tag{6.49}$$

obtenue à partir de (6.42). L'équation (6.49) étant à variables séparées, elle peut être intégrée exactement. On obtient

$$\alpha u + v - \ln u^{\alpha} = H \tag{6.50}$$

où H désigne une constante d'intégration. La valeur minimale de H compatible avec $u,v \geq 0$ est $H_{min=}1+\alpha$. La trajectoire de phase correspondant à ce minimum se réduit au seul point d'équilibre stable $(1,1)$. À toutes les valeurs de $H > H_{min}$ correspond une trajectoire de phase fermée qui tourne autour du centre $(1,1)$ (Cf. Fig. 6.1). Ceci confirme et étend les résultats obtenus au voisinage des points critiques.

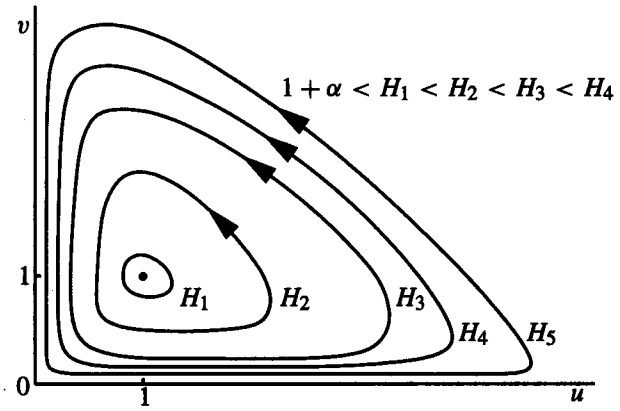


FIG. 6.1 – Plan de phase

Une trajectoire fermée dans le plan de phase représente une solution périodique présentant des oscillations (pas nécessairement harmoniques) des deux variables. La condition initiale détermine l'amplitude et la période des oscillations. Une solution type est présentée à la figure 6.2. Les oscillations des deux populations sont déphasées d'environ un quart de période ; la population de la proie est maximale ou minimale lorsque la population du prédateur est égale à sa valeur d'équilibre, elle est égale à sa valeur d'équilibre lorsque la population du prédateur est extrémale. Les maxima des populations du prédateur suivent immédiatement les maxima de la proie.

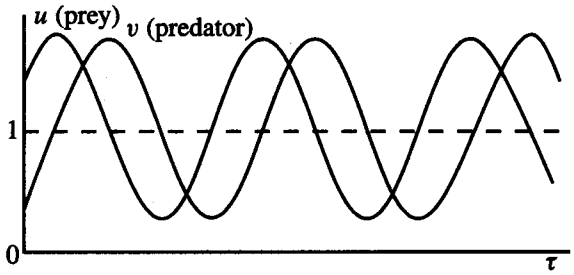


FIG. 6.2 – Solution type du modèle.

La relation (6.50) constitue une *intégrale première* du système, *i.e.* une relation entre les variables qui est conservée au cours du temps. Un système possédant une telle intégrale première est qualifié de *conservatif*.

Le modèle de Lotka-Volterra constitue un exemple très utilisé d'interaction entre espèces. Malheureusement, il souffre d'un problème structural qui limite son applications

aux systèmes réels. En effet, si on perturbe légèrement le système décrivant une trajectoire donnée, celui-ci se retrouve sur une trajectoire différente caractérisée par un amplitude et une période modifiées. La nouvelle trajectoire n'est pas partout très proche de la trajectoire initiale. Ce problème dit d'*instabilité structurale* est classique des systèmes conservatifs.

Pour la petite histoire, citons l'application de Gilpin (1973) du modèle de Lotka-Volterra à la modélisation de la dynamique des populations de lièvre et de lynx, les premiers constituant évidemment la proie des seconds. Le modèle de Lotka-Volterra fournit des trajectoires de phases proches de celles déduites des relevés de capture d'animaux à fourrure dans la Baie d'Hudson entre 1845 et 1835, les populations des deux espèces fluctuant dans le temps. Malheureusement, les données montrent que les maxima de population de lynx précèdent les maxima de population de lièvre suggérant donc que les lièvres se nourrissent de lynx ! L'explication la plus probable de ce paradoxe est certainement à chercher du côté de la stratégie des trappeurs qui devaient se détourner de la capture des lièvres lorsque les populations étaient faibles et se tournaient alors vers la capture plus profitable du lynx. Les données de capture ne donnent donc pas une image fiable des variations des populations vraies des deux espèces.

Des variantes du modèle de Lotka-Volterra ont été proposées. Ainsi, si on représente la croissance de la proie par un modèle logistique, on a

$$\frac{dN}{dt} = a(N - N/K) - bN\,P, \qquad \frac{dP}{dt} = cN\,P - dP \tag{6.51}$$

De même, on peut supposer qu'une population N_0 de la proie peut échapper au prédateur. Ceci peut se produire si la dispersion géographique des proies est telle que la probabilité de rencontre entre la proie et le prédateur est faible. Le prédateur dépense alors trop d'énergie pour couvrir l'ensemble du territoire et est incapable d'attraper toutes les proies. Dans ce cas, on écrira

$$\frac{dN}{dt} = a(N - N/K) - b(N - N_0)\,P, \qquad \frac{dP}{dt} = c(N - N_0)\,P - dP \tag{6.52}$$

Ces modèles peuvent être analysés en utilisant les mêmes techniques à celles utilisées ci-dessus.

6.2 Analyse dans l'espace de phase.

Ayant introduit le plan de phase dans l'exemple de Lotka-Volterra, étudions de façon systématique les différents types de comportement que nous pouvons rencontrer dans le plan de phase. Pour ce faire, nous considérons un modèle général à deux équations de la forme

$$\dot{x} = f(x,y), \qquad \dot{y} = g(x,y) \tag{6.53}$$

où f et g sont des fonctions connues et où le point surmontant une variable désigne la dérivée de celle-ci par rapport au temps.

Le théorème donné en annexe du chapitre précédent détermine des conditions générales d'existence et d'unicité de la solution de (6.53).

Tout point (x_0, y_0) tel que

$$0 = f(x_0, y_0), \qquad 0 = g(x_0, y_0) \tag{6.54}$$

est appelé un *point critique, point fixe* ou *point d'équilibre* du système. Si les conditions initiales du système sont prises en un tel point, le système ne quitte jamais cet état.

La courbe décrite de façon paramétrique par $(x(t), y(t))$ dans le plan de phase (x, y) constitue la trajectoire de phase du système. Le système (6.53) possède une et une seule solution correspondant à chaque point du plan de phase au voisinage duquel les fonctions f et g vérifient les conditions du théorème d'existence de la solution. Dans le cas d'un système *autonome* comme (6.42), *i.e.* d'un système dont la dynamique ne dépend pas explicitement du temps, chaque tel point du plan de phase se trouve sur une et une seule trajectoire. En d'autres termes, les trajectoires ne peuvent se croiser, sauf éventuellement en certains points particuliers appelés *points singuliers.*

La direction d'évolution du système lorsque le temps t augmente peut être indiquée par une flèche attachée à chaque trajectoire.

Le pente de la trajectoire en chacun des points du plan de phase est donnée par

$$\frac{dy}{dx} = \frac{\dot{y}}{\dot{x}} = \frac{g(x,y)}{f(x,y)} \tag{6.55}$$

en chacun des points pour lesquels le second membre de cette équation est bien défini (ou même infini). La trajectoire est verticale en un point (x, y) où f s'annule et g diffère de zéro. Elle est horizontale là où $g(x, y) = 0$ mais $f(x, y) \neq 0$. Les points singuliers sont ceux qui laissent le second membre indéterminé. En particulier, les trajectoires correspondant à un point d'équilibre stable se réduisent à un seul point.

Toute évolution périodique du système est représentée dans le plan de phase par une trajectoire fermée.

6.2.1 Stabilité locale

La stabilité locale des points critiques (x_0, y_0) du système (6.53) peut être étudiée en linéarisant les équations au voisinage de ce point critique. Introduisons les perturbations ξ, η telles que

$$x = x_0 + \xi, \qquad y = y_0 + \eta \tag{6.56}$$

Si ξ et η représentent de faibles perturbations de (x_0, y_0), on peut approcher f par les premiers termes de son développement de Taylor, soit, en tenant compte de (6.54),

$$f \approx \left(\frac{\partial f}{\partial x}\right)_0 \xi + \left(\frac{\partial f}{\partial y}\right)_0 \eta \tag{6.57}$$

où la notation $(\cdot)_0$ signifie que les dérivées sont évaluées au point (x_0, y_0). Faisant de même pour g, l'approximation de (6.53) s'écrit

$$\dot{\xi} = a\,\xi + b\,\eta, \qquad \dot{\eta} = c\,\xi + d\,\eta \tag{6.58}$$

où on a introduit les constantes

$$a = \left(\frac{\partial f}{\partial x}\right)_0, \quad b = \left(\frac{\partial f}{\partial y}\right)_0, \quad c = \left(\frac{\partial g}{\partial x}\right)_0, \quad d = \left(\frac{\partial g}{\partial y}\right)_0 \tag{6.59}$$

De façon équivalente, le système linéaire, homogène, à coefficients constants (6.58) peut s'écrire sous forme matricielle

$$\frac{d}{dt}\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix} \tag{6.60}$$

Toute l'information sur la dynamique du système et l'interaction entre les espèces qui le compose est contenue dans la matrice de communauté ('community matrix')

$$\mathsf{A} = \begin{pmatrix}a & b\\ c & d\end{pmatrix} \tag{6.61}$$

La solution de (6.60) peut généralement être exprimée en fonction des valeurs propres et des vecteurs propres de A. A cet effet,recherchons des solutions de (6.60) de la forme particulière

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\xi_0\\ \eta_0\end{pmatrix}\mathrm{e}^{\lambda t} \tag{6.62}$$

i.e. des solutions où les deux variables indépendantes varient proportionnellement. Substituant cette expression dans (6.60), on voit qu'une telle solution n'existe que si

$$\mathsf{A}\begin{pmatrix}\xi_0\\ \eta_0\end{pmatrix} = \lambda\begin{pmatrix}\xi_0\\ \eta_0\end{pmatrix} \tag{6.63}$$

i.e. si λ est une valeur propre de A. Celles-ci sont obtenue en résolvant l'équation caractéristique

$$\mathrm{dtm}\,(\mathsf{A} - \lambda\mathbb{I}) = 0 = \begin{vmatrix}a-\lambda & b\\ c & d-\lambda\end{vmatrix} \tag{6.64}$$

où $\mathbb{I}$ désigne la matrice identité. Dans les cas les plus simples, lorsque les valeurs propres de A sont distinctes, la solution complète du problème linéarisé (6.60) s'écrit comme la combinaison des modes fondamentaux

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix}(t) = C_1\mathsf{v}^{(1)}\begin{pmatrix}\xi^{(1)}\\ \eta^{(1)}\end{pmatrix}\mathrm{e}^{\lambda_1 t} + C_2\begin{pmatrix}\xi^{(2)}\\ \eta^{(2)}\end{pmatrix}\mathsf{v}^{(2)}\,\mathrm{e}^{\lambda_2 t} \tag{6.65}$$

où C_1 et C_2 sont des constantes d'intégration et où

$$\mathsf{v}^{(1)} = \begin{pmatrix}\xi^{(1)}\\ \eta^{(1)}\end{pmatrix}, \qquad \mathsf{v}^{(2)} = \begin{pmatrix}\xi^{(2)}\\ \eta^{(2)}\end{pmatrix} \tag{6.66}$$

sont les vecteurs propres associés à λ_1 et à λ_2.

On en déduit que l'équilibre est (asymptotiquement) stable si la partie réelle de toutes les valeurs propre est strictement négative. Il est instable si une valeur propre (ou plus) présente une partie réelle strictement positive. Si toutes les valeurs propres possèdent une partie réelle négative mais que certaines sont purement imaginaires, l'équilibre est marginalement stable ou faiblement instable[1] : la solution présente des oscillations non amorties (marginalement stable) ou dont l'amplitude croît de façon polynomiale (faiblement instable). Il faut cependant être attentif au fait que ces conclusions peuvent être mises à mal par les termes non linéaires négligés. Ce sont eux qui, dans le cas de la stabilité marginale et de l'instabilité faible, commandent le caractère stable ou instable.

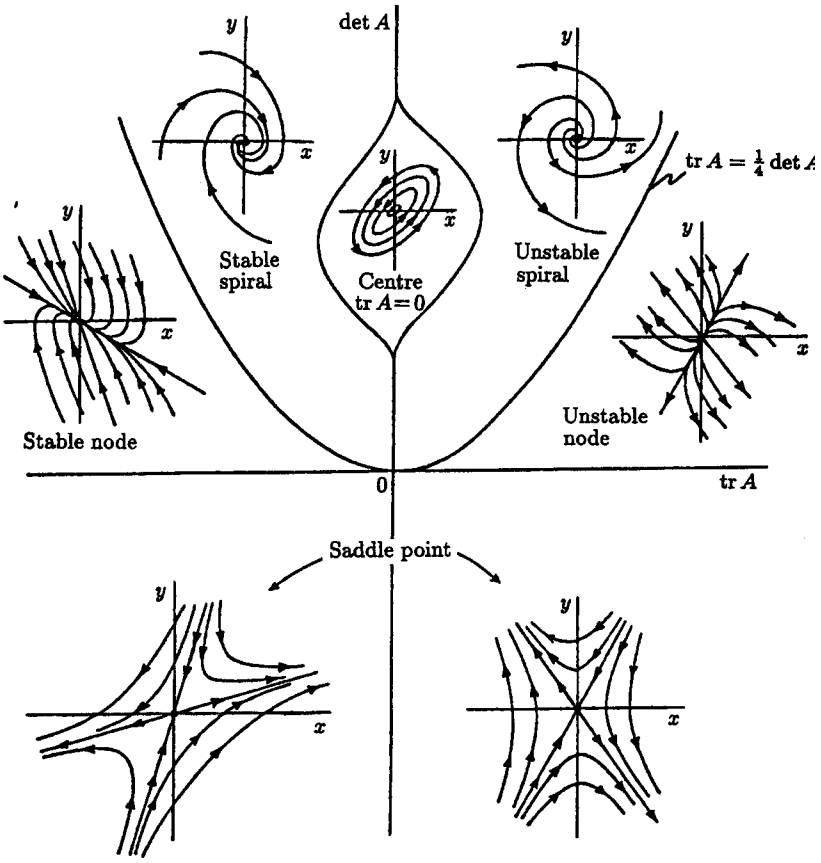


FIG. 6.3 – Représentation dans le plan de phase de la dynamique au voisinage des différents types de points critiques.

Les vecteurs propres correspondant aux valeurs propres λ_1, λ_2 représentent les directions de l'espace de phase selon lesquelles on observe la croissance ou la

[1] Pour être précis, l'équilibre est marginalement stable si les blocs de Jordan associés aux valeurs propres purement imaginaires sont d'ordre un. Il est faiblement instable sinon.

décroissance exponentielle correspondant à la stabilité ou à l'instabilité de la solution. En fonction des valeurs de λ_1 et λ_2, on peut donc distinguer les cas suivants :

i. λ_1 et λ_2 réelles et distinctes :
   (a) $\lambda_2 < \lambda_1 < 0$. Les trajectoires de phase convergent vers le point d'équilibre stable en s'alignant asymptotiquement avec $v^{(1)}$ (la décroissance dans la direction de $v^{(2)}$ est plus rapide). Le point d'équilibre est un *nœud stable.*
   (b) $0 < \lambda_2 < \lambda_1$. Les trajectoires de phase divergent du point d'équilibre instable en s'alignant asymptotiquement avec le vecteur propre relatif à la valeur propre la plus grande. On a un *nœud instable.*
   (c) $\lambda_2 < 0 < \lambda_1$. Les trajectoires divergent selon $v^{(1)}$ et convergent vers le point d'équilibre selon $v^{(2)}$. Le point d'équilibre instable est un *point de selle.*
   Remarquons que, si les conditions initiales sont alignées avec $v^{(2)}$, le système tend vers sa configuration d'équilibre instable. Ce point ne peut cependant pas être atteint en un temps fini. On parle de *mouvement asymptotique* vers la position d'équilibre.

ii. λ_1 et λ_2 complexes (nécessairement conjugués), *i.e.* $\lambda_1, \lambda_2 = \alpha \pm i\beta$ avec $\beta \neq 0$.
   (a) $\alpha < 0$. L'équilibre est stable et les trajectoires de phase s'enroulent en convergeant vers le point d'équilibre qui est qualifié de *foyer stable.*
   (b) $\alpha = 0$. Les trajectoires de phase sont des courbes fermées tournant autour du point d'équilibre. Celui-ci est un *centre.* Les trajectoires correspondantes sont périodiques.
   (c) $\alpha > 0$. Les trajectoires de phase divergent en tournant autour du point d'équilibre instable. On parle de *foyer instable.*

iii. $\lambda_1 = \lambda_2$. Si les valeurs propres sont confondues, le point d'équilibre apparaît comme un nœud (stable ou instable en fonction du signe des valeurs propres) éventuellement *dégénéré* [2]. Dans ce dernier cas, les trajectoires sont radiales.

6.2.2 Stabilité globale, solutions périodiques et cycles limites.

L'analyse linéaire de stabilité exposée dans la section précédente n'est valable que localement. Lorsqu'un point d'équilibre est présenté comme asymptotiquement stable par linéarisation, cela signifie que les perturbations suffisamment petites de l'équilibre donnent naissance à des trajectoires qui tendent vers le point d'équilibre stable. À partir d'une certaine ampleur des perturbations, les trajectoires peuvent être qualitativement différentes et, par exemple, s'écarter du point d'équilibre stable étudié. En général, la région du plan de phase dont sont issues les trajectoires convergeant effectivement vers un point d'équilibre stable est appelée le *basin d'attraction* de ce point. Un point d'équilibre est qualifié de *globalement stable* lorsque son bassin d'attraction couvre l'ensemble du plan de phase. Dans ce cas, un seul équilibre stable peut évidemment exister.

[2] On a une étoile si les blocs de Jordan sont tous de taille unitaire et un nœud sinon.

L'analyse linéaire ne permet pas de déterminer la taille de ce bassin d'attraction. La connaissance de celle-ci est cependant capitale pour déterminer l'ampleur des perturbations auxquelles un système peut faire face sans modification qualitative de son comportement.

Le plan de phase est généralement séparés en différents bassins d'attraction correspondant à différents points d'équilibre stable. Les limites des bassins d'attraction sont matérialisées par des trajectoires particulières appelées *séparatrices*. Celles-ci peut parfois prendre la forme de courbes fermées correspondant à des trajectoires périodiques ou *cycles*. L'existence de telles trajectoires périodiques (sans résoudre complètement le système d'équations différentielles) peut parfois être déduites de l'application de théorèmes appropriés. De même, l'existence de trajectoires périodiques peut aider à la découverte de points critiques.

On montre que, si un système possède une solution périodique représentée par une courbe fermée simple $\mathcal{C}$ dans le plan de phase, alors celle-ci contient au moins un point critique. Si $\mathcal{C}$ entoure un seul point critique, celui-ci ne peut être un point de selle. Inversement, si une région simplement connexe Ω du plan de phase d'un système autonome ne contient aucun point critique ou un seul point de selle, alors Ω ne peut contenir aucune solution périodique.

Le *théorème de Poincarré-Bendixson* repose sur la notion de *région invariante* pour prédire l'existence de solutions périodiques. Une région invariante Ω est une région du plan telle que toutes les trajectoires générées à partir de conditions initiales choisies dans Ω sont entièrement contenues dans Ω à tous les instants ultérieurs. Un point d'équilibre, une trajectoire périodique ou le basin d'attraction d'un point d'équilibre stable constituent autant de cas particuliers de régions invariantes. De telles régions invariantes peuvent être identifiées en observant que les trajectoires ne peuvent que rentrer mais jamais sortir Ω. Soit donc une région invariante Ω du plan ne contenant aucun point critique sur sa frontière :

i. si Ω est connexe de type I, *i.e.* si Ω est une région bornée délimitée par une courbe simple, et contient un nœud instable unique ou un foyer instable unique, il existe au moins une solution périodique dans Ω.
ii. si Ω est connexe de type II, *i.e.* si Ω est une région annulaire bornée comprise entre deux courbes simples, et ne contient aucun point critique, il existe au moins une trajectoire périodique dans Ω.

Dans les deux cas, toute trajectoire non périodique contenue dans Ω spirale en convergeant vers la solution périodique qui est dès lors appelée *cycle limite*. L'ensemble des points conduisant à un cycle limite définit le bassin d'attraction de ce cycle limite.

6.2.3 Généralisation.

Les notions précédentes peuvent être généralisées aux systèmes de n équations différentielles du premier ordre. Les concepts de stabilité ne sont pas modifiés. De même, une solution décrit toujours une trajectoire qui est maintenant une courbe dans un espace

à n dimensions. Cet espace, qui généralise la notion de plan de phase, est appelé l'*espace de phase*. Les représentations graphiques d'un tel espace sont évidemment impossibles pour $n > 3$. Par contre, on peut considérer des représentations graphiques des coupes bidimensionnelles de cet espace de phase.

Tout système comportant des équations d'un ordre supérieur au premier ordre peut être transformé en un système équivalent du premier ordre en introduisant des variables dépendantes supplémentaires. Par exemple, l'équation du troisième ordre

$$x^{(3)} + a_2(x,y)\ddot{x} + a_1(x,y)\dot{x} + a_0(x,y) = f(x,y) \tag{6.67}$$

est équivalente à

$$\begin{cases} \dot{x} &= \tilde{x}_2 \\ \dot{\tilde{x}}_2 &= \tilde{x}_3 \\ \dot{\tilde{x}}_3 &= f(x,y) - a_2(x,y)\tilde{x}_2 - a_1(x,y)\tilde{x}_1 - a_0(x,y)x \end{cases} \tag{6.68}$$

Enfin, l'analyse dans l'espace de phase peut aussi être étendue aux systèmes dépendants du temps en introduisant une variable dépendante supplémentaire qui s'identifie au temps. Ainsi, on remplacera le système

$$\dot{x} = f(t,x,y), \qquad \dot{y} = g(t,x,y) \tag{6.69}$$

par

$$\dot{x} = f(z,x,y), \qquad \dot{y} = g(z,x,y), \qquad \dot{z} = 1 \tag{6.70}$$

Un système de n équations du premier ordre dépendant explicitement du temps est donc équivalent à un système autonome de dimension $n+1$ dont la dynamique peut être analysée dans le plan de phase.

Les notions de noeud, centre, point de selle, bassin d'attraction, cycle limites,... se généralisent également (mais pas le théorème de Poincarré-Bendixson). L'augmentation de la dimension de l'espace de phase permet cependant une plus grande variété dans les types de solutions. En particulier, à partir de la dimension 3, des solutions peuvent être bornées sans donner naissance à des cycle limites ou points d'équilibre : les trajectoires pouvant se tordre, se contourner, s'enchevêtrer en de fantastiques nœuds connus sous le nom d'*attracteurs étranges*. Ceux-ci permettent à deux solutions initialement très voisines de diverger rapidement tout en restant dans un région bornée de l'espace de phase : c'est le régime du chaos.

6.2.4 Compétition et symbiose.

Le modèle de Lotka-Volterra décrit la relation de prédateur-proie au moyen de deux équations différentielles couplées. De la même façon, on peut aisément décrire la dynamique couplée de deux espèces en compétition pour les mêmes ressources ou la symbiose de deux espèces, *i.e.* lorsque la cohabitation de deux espèces se réalise au bénéfice de ces espèces.

Compétition.

Considérons tout d'abord le cas de deux espèces en compétition. Notons N_1 et N_2 les populations correspondantes. Si on suppose que les deux populations sont caractérisées par une croissance logistique en l'absence de l'autre espèce, on écrira, par exemple

$$\begin{cases} \dfrac{dN_1}{dt} = r_1 N_1 \Big[1 - \dfrac{N_1}{K_1} - b_{12}\dfrac{N_2}{K_1}\Big] \\ \dfrac{dN_2}{dt} = r_2 N_2 \Big[1 - \dfrac{N_2}{K_2} - b_{21}\dfrac{N_1}{K_2}\Big] \end{cases} \tag{6.71}$$

où r_1, r_2, K_1, K_2, b_{12} et b_{21} sont des constantes positives. Les r_i désignent les taux de croissance linéaire et les K_i les capacités portantes. Les constantes b_{12} et b_{12} mesurent l'influence réciproque des deux espèces. Les termes correspondant montrent la réduction du taux de croissance d'une espèce induite par la présence de l'autre espèce.

Les équations (6.71) peuvent être rendues adimensionnelles en posant

$$\begin{gathered} u_1 = \frac{N_1}{K_1}, \qquad u_2 = \frac{N_2}{K_2}, \qquad \tau = r_1 t, \qquad \rho = \frac{r_2}{r_1} \\ a_{12} = b_{12}\frac{K_2}{K_1}, \qquad a_{21} = b_{21}\frac{K_1}{K_2} \end{gathered} \tag{6.72}$$

Il vient

$$\begin{cases} \dfrac{du_1}{dt} = u_1(1 - u_1 - a_{12}u_2) = f_1(u_1, u_2) \\ \dfrac{du_2}{dt} = \rho u_2(1 - u_2 - a_{21}u_1) = f_2(u_1, u_2) \end{cases} \tag{6.73}$$

Les points critiques sont les solutions de $f_1(u_1^*, u_2^*) = f_2(u_1^*, u_2^*) = 0$, soit les couples (u_1^*, u_2^*) donnés par

$$(0,0), \qquad (1,0), \qquad (0,1), \qquad (\tilde{u}_1, \tilde{u}_2) = \left(\frac{1 - a_{12}}{1 - a_{12}a_{21}}, \frac{1 - a_{21}}{1 - a_{12}a_{21}}\right) \tag{6.74}$$

(où le dernier point critique n'est possible que pour certaines gammes de valeurs des paramètres).

La stabilité des points critiques dépend des valeurs propres de la matrice de communauté

$$\mathsf{A} = \begin{pmatrix} \dfrac{\partial f_1}{\partial u_1} & \dfrac{\partial f_1}{\partial u_2} \\ \dfrac{\partial f_2}{\partial u_1} & \dfrac{\partial f_2}{\partial u_2} \end{pmatrix}_{(u_1^*, u_2^*)} = \begin{pmatrix} 1 - 2u_1^* - a_{12}u_2^* & -a_{12}u_1^* \\ -\rho a_{21}u_2^* & \rho(1 - 2u_2^* - a_{12}u_1^*) \end{pmatrix} \tag{6.75}$$

On vérifie aisément que le point $(0,0)$ est toujours instable puisque la matrice de communauté correspondante s'écrit

$$\mathsf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \tag{6.76}$$

dont les valeurs propres 1 et ρ sont strictement positives.

La configuration $(1,0)$ donne lieu à la matrice de communauté

$$A = \begin{pmatrix} -1 & -a_{12} \\ 0 & \rho(1-a_{21}) \end{pmatrix} \tag{6.77}$$

dont les valeurs propres sont -1 et $\rho(1-a_{21})$. Par conséquent, l'équilibre est stable[3] pour $a_{21} > 1$ et instable pour $a_{21} < 1$.

De même, la matrice de communauté relative à l'équilibre $(0,1)$ s'écrit

$$A = \begin{pmatrix} 1-a_{12} & 0 \\ -\rho & -\rho \end{pmatrix} \tag{6.78}$$

dont les valeurs propres sont $1-a_{12}$ et $-\rho$. Par conséquent, l'équilibre est stable pour $a_{12} > 1$ et instable pour $a_{12} < 1$.

Lorsque les conditions sont remplies pour que le troisième équilibre de (6.74) soit possible, sa stabilité dépend du signe (de la partie réelle) des valeurs propres de

$$A = \frac{1}{1-a_{12}a_{21}} \begin{pmatrix} a_{12}-1 & a_{12}(a_{12}-1) \\ \rho a_{21}(a_{21}-1) & \rho(a_{21}-1) \end{pmatrix} \tag{6.79}$$

soit

$$\lambda_{1,2} = \frac{1}{2(1-a_{12}a_{21})} \Big[(a_{12}-1) + \rho(a_{21}-1) \pm \sqrt{[(a_{12}-1)+\rho(a_{21}-1)]^2 - 4\rho(1-a_{12}a_{21})(a_{12}-1)(a_{21}-1)} \Big] \tag{6.80}$$

Sans entrer dans la discussion du signe de (6.80), on peut remarquer que le système est globalement stable. En effet, les trajectoires de phases pointent toutes vers l'intérieur du rectangle $[0,U_1] \times [0,U_2]$ si on prend U_1 et U_2 suffisamment grands.

Considérons maintenant les différentes combinaisons possibles des paramètres du problème.

a) $a_{12} < 1, a_{21} < 1$.
   Les points $(1,0)$ et $(0,1)$ sont tous deux instables et constituent des points de selle. L'équilibre est stable au point $(\tilde{u}_1, \tilde{u}_2)$. Toutes les trajectoires convergent vers ce point (centre).

b) $a_{12} > 1, a_{21} > 1$.
   Pour de telles valeurs des paramètres, les configurations (1,0) et (0,1) sont toutes les deux stables. Puisque $1-a_{12}a_{21} < 0$, le point $(\tilde{u}_1, \tilde{u}_2)$ est aussi admissible et

[3] Le cas $a_{21} = 1$ est donne lieu à un équilibre marginalement stable au sens de l'analyse infinitésimale. Une analyse non linéaire complète s'impose alors.

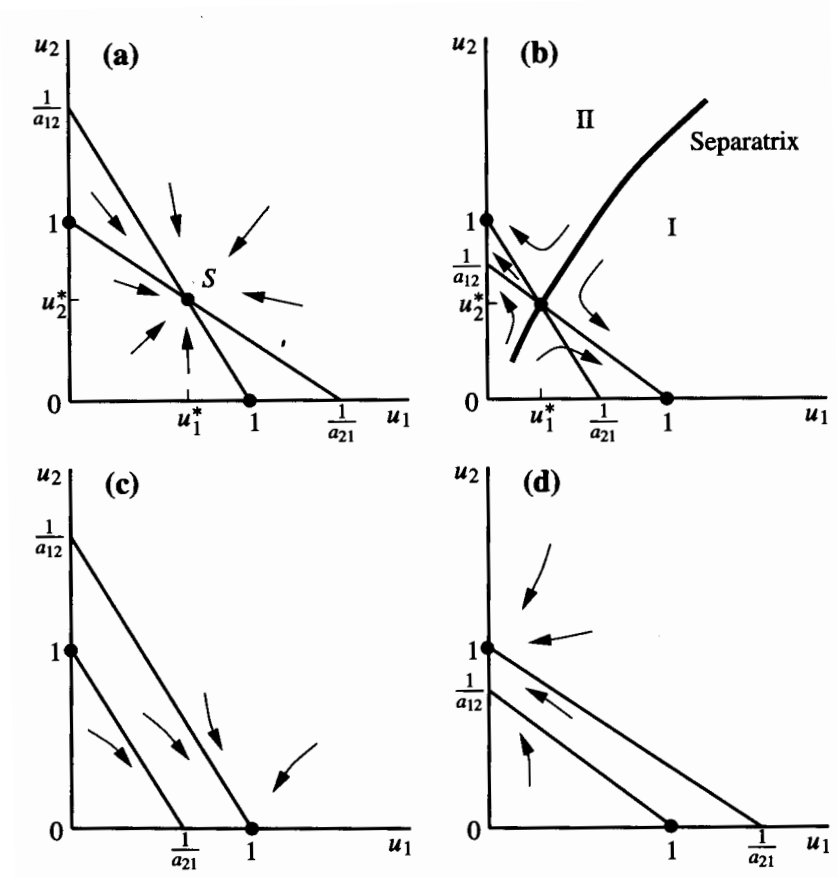


FIG. 6.4 – Compétition entre espèces : plan de phase

les valeurs propres de la matrice de communauté correspondante sont de signes différents. [4]

L'équilibre $(\tilde{u}_1, \tilde{u}_2)$ est donc instable : il correspond à un point de selle.

Le plan de phase peut être esquissé en raisonnant sur les isoclines de f_1 et f_2 définissant les signes de

$$\frac{du_1}{dz} \qquad \text{et} \qquad \frac{du_2}{dz}$$

Il fait apparaître une trajectoire particulière passant par le point de selle $(\tilde{u}_1, \tilde{u}_2)$ et séparant le plan de phase en deux régions disjointes correspondant aux deux bassins d'attraction des points d'équilibre stable : toute trajectoire issue d'un point de la région I (resp. II) tend asymptotiquement vers le point d'équilibre stable (1,0) (resp. (0,1)).

c) et d) $a_{12} < 1, a_{21} > 1$ ou $a_{12} > 1$ et $a_{21} < 1$.

Les points $(1,0)$ et $(0,1)$ sont les deux seuls points d'équilibre. L'un est stable et l'autre instable.

Lorsque $a_{12}, a_{21} > 1$ ou lorsque $a_{12} > 1$ et $a_{21} < 1$ ou bien encore quand $a_{12} < 1$ et $a_{21} > 1$, l'une des deux espèces est mieux armée que l'autre dans la compétition et finit par l'emporter, quelles que soient les conditions initiales. Le système se stabilise alors par la disparition complète d'une des deux espèces. C'est le principe de l'exclusion compétitive, connue aussi sous le nom de *principe de Gause*, qui affirme que deux espèces ne peuvent occuper durablement la même niche écologique ; si deux espèces sont en compétition pour une même ressource essentielle, l'une fera mieux que l'autre.

Le cas $a_{12} < 1$ et $a_{21} < 1$ est le seul cas où une cohabitation stable des deux espèces est possible. Le niveau d'équilibre de cette cohabitation est, évidemment, inférieur au niveau d'équilibre des espèces considérées individuellement. On remarque que cette cohabitation demande

$$b_{12}\frac{K_2}{K_1} < 1 \tag{6.81}$$

ce qui pourrait se traduire par le fait que la compétition entre les deux espèces ne doit pas être trop sévère.

[4] Le polynôme caractéristique peut s'écrire sous la forme

$$\begin{aligned}(\lambda-\lambda_1)(\lambda-\lambda_2) &= \lambda^2-(\lambda_1+\lambda_2)\lambda+\lambda_1\lambda_2 \\ &= \lambda^2-\text{trace A}\ \lambda+\det \mathsf{A}\end{aligned}$$

où on a noté λ_1 et λ_2 les valeurs propres de A et où

$$\text{trace A} = \frac{a_{12}-1+\rho(a_{21}-1)}{1-a_{12}a_{21}} < 0$$

$$\det \mathsf{A} = \rho\frac{(a_{12}-1)(a_{21}-1)}{1-a_{12}a_{21}} < 0$$

Si b_{12} et b_{21} sont grands et que K_1 et K_2 sont semblables, alors les deux espèces se livrent à une concurrence féroce et l'équilibre stable ne peut être réalisé que par la disparition complète d'une des deux espèces ($a_{12} > 1$ et $a_{21} > 1$). La détermination de l'espèce qui l'emporte dépend de l'avantage initial de l'une ou l'autre espèce.

Il convient de remarquer que ce sont les groupements adimensionnels a_{12} et a_{21} (et éventuellement les conditions initiales) qui gouvernent l'évolution ultime du système. Le ratio ρ des taux de croissance n'affecte en rien la stabilité des solutions. Il modifie seulement l'échelle de temps de la dynamique du système.

6.2.5 Mutualisme ou symbiose.

Le modèle de Lotka-Volterra peut également être adapté pour décrire la dynamique d'un système où l'interaction de deux ou plusieurs espèces se produit à l'avantage de toutes ces espèces, *i.e.* où les espèces vivent en symbiose. Pour écrire un modèle réaliste de se genre de comportement, il est évidemment nécessaire d'inclure une limitation de la croissance, par exemple par la disponibilité limitée des ressources communes. Un modèle simple de ce type est donné par

$$\begin{cases} \dfrac{dN_1}{dt} &= r_1 N_1 \left(1 - \dfrac{N_1}{K_1} + b_{12}\dfrac{N_2}{K_1}\right) \\ \dfrac{dN_2}{dt} &= r_2 N_2 \left(1 - \dfrac{N_2}{K_2} + b_{21}\dfrac{N_1}{K_2}\right) \end{cases} \tag{6.82}$$

où r_1 et r_2 sont les taux de croissance linéaire des deux espèces, où K_1 et K_2 les capacités de l'environnement pour ces deux espèces et où $b_{12} > 0$ et $b_{21} > 0$ caractérisent l'augmentation du taux de croissance d'une espèce en présence de l'autre.

Le système (6.82) peut être écrit en variables adimensionnelles en posant,

$$\begin{gathered} u_1 = \frac{N_1}{K_1}, \qquad u_2 = \frac{N_2}{K_2}, \qquad \tau = r_1 t, \qquad \rho = \frac{r_2}{r_1} \\ a_{12} = b_{12}\frac{K_2}{K_1}, \qquad a_{21} = b_{12}\frac{K_1}{K_2} \end{gathered} \tag{6.83}$$

Il vient

$$\begin{cases} \dfrac{du_1}{dt} = u_1(1 - u_1 + a_{12}u_2) = f_1(u_1, u_2) \\ \dfrac{du_2}{dt} = \rho u_2(1 - u_2 + a_{21}u_1) = f_2(u_1, u_2) \end{cases} \tag{6.84}$$

Les points fixes sont

$$\begin{gathered} (0,0), \qquad (1,0), \qquad (0,1) \\ (\tilde{u}_1, \tilde{u}_2) = \left(\frac{1 + a_{12}}{1 - a_{12}a_{21}}, \frac{1 + a_{22}}{1 - a_{12}a_{21}}\right) \qquad \text{si} \qquad 1 - a_{12}a_{21} > 0 \end{gathered} \tag{6.85}$$

En procédant comme à la section précédente (seuls les signes des coefficients a_{12} et a_{21} sont modifiés), on montre aisément que $(0,0)$ est un noeud instable tandis que les points $(1,0)$ et $(0,1)$ sont des points de selle. Ces trois points sont donc instables. Si $1-a_{12}a_{21}<0$, le système ne peut donc présenter d'équilibre stable ; les deux populations explosent (sont non bornées).

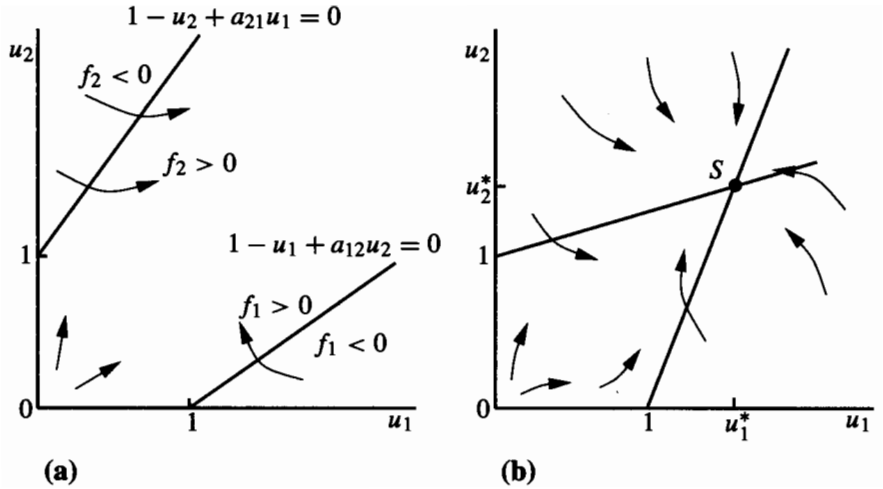


FIG. 6.5 – Plan de phase dans le cas où a) $1-a_{12}a_{21}<0$ et b) $1-a_{12}a_{21}>0$

Si $1-a_{12}a_{21}>0$, l'équilibre est stable en $(\tilde{u}_1,\tilde{u}_2)$: toutes les trajectoires convergent vers ce noeud stable. Remarquons que cet équilibre est caractérisé par $\tilde{u}_1,\tilde{u}_2>1$, ce qui correspond à $N_1>K_1$, $N_2>K_2$. L'interaction des deux espèces permet donc le maintien de populations supérieures aux populations d'équilibre lorsque les espèces vivent isolément.

6.3 Modèles discrets pour l'interaction des populations.

Les modèles d'interaction présentés dans la section précédente peuvent être transposés au cas discret. Considérons en particulier le cas du système proie-prédateur. Désignant la proie par N_k et le prédateur par P_k, on écrira, par exemple,

$$\begin{cases} N_{k+1} = r\,N_k\,\exp(-a\,P_k) \\ \\ P_{k+1} = N_k\,[1-\exp(-a\,P_k)] \end{cases} \tag{6.86}$$

où $r(>0)$ représente le taux de croissance net de la proie en l'absence de prédation et où $a(>0)$ mesure l'intensité de l'interaction entre les deux espèces.

Le système possède deux états d'équilibre

$$\begin{cases} N^* = 0 \\ \\ P^* = 0 \end{cases} \quad \text{ou} \quad \begin{cases} P^* = \dfrac{\ln r}{a} \\ \\ N^* = \dfrac{r \ln r}{a(r-1)} \end{cases} \tag{6.87}$$

Le second équilibre n'existe que pour $r > 1$.

La stabilité linéaire de ces équilibres peut être étudiée, comme d'habitude, par linéarisation des équations en écrivant

$$\begin{cases} N_k = N^* + \varepsilon_k \\ \\ P_k = P^* + \eta_k \end{cases} \quad \text{où} \quad \left|\frac{\varepsilon_k}{N^*}\right| \ll 1, \quad \left|\frac{\eta_k}{N^*}\right| \ll 1 \tag{6.88}$$

Dans le cas de l'équilibre trivial $N^* = P^* = 0$, il vient

$$\varepsilon_{k+1} = r\,\varepsilon_k; \qquad \eta_{k+1} = 0 \tag{6.89}$$

L'équilibre est donc stable si $r < 1$ ($N_k \to 0$ pour $k \to +\infty$), instable pour $r > 1$ et marginalement pour $r = 1$ ($N_k = N_0\ \forall k$).

Pour $r > 1$, les perturbations de la seconde configuration d'équilibre évoluent selon

$$\begin{cases} \varepsilon_{k+1} = \varepsilon_k - N^* a\,\eta_k \\ \\ \eta_{k+1} = \varepsilon_k \left(1 - \dfrac{1}{r}\right) + \dfrac{N^* a}{r}\eta_k \end{cases} \tag{6.90}$$

La solution analytique de ce système linéaire peut être obtenue en itérant la première équation et en éliminant η_k et η_{k+1} au profit de ε_k et ε_{k+1} pour se ramener à une équation du second ordre en ε_k :

$$\varepsilon_{k+2} - \left(1 - \frac{N^* a}{r}\right)\varepsilon_{k+1} + N^* a \varepsilon_k = 0 \tag{6.91}$$

Cette équation peut ensuite être résolue par la méthode du polynôme caractéristique.

De façon équivalente, on peut transposer la méthode matricielle de résolution introduite dans le cadre des systèmes d'équations différentielles linéaires. On a

$$\begin{pmatrix} \varepsilon_{k+1} \\ \eta_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & -N^* a \\ \\ 1 - \dfrac{1}{r} & \dfrac{N^* a}{r} \end{pmatrix} \begin{pmatrix} \varepsilon_k \\ \eta_k \end{pmatrix} \tag{6.92}$$

ce que l'on peut noter

$$\mathsf{x}_{k+1} = \mathsf{A}\mathsf{x}_k \tag{6.93}$$

Si on recherche des solutions de la forme

$$\mathsf{x}_k = \lambda^k \mathsf{x}_0, \tag{6.94}$$

il vient, après substitution dans (6.93),

$$\mathsf{A}\mathsf{x}_0 = \lambda \mathsf{x}_0 \tag{6.95}$$

i.e. λ est une valeur propre de A et x_0 un vecteur propre associé. Le système sera stable si toutes les valeurs propres de A sont inférieures à 1 en module. Si une des valeurs propres est telle que $|\lambda| > 1$, le mode correspondant croît exponentiellement et le système est instable.

Dans le cas particulier étudié, on a

$$|\mathsf{A} - \lambda I| = \begin{vmatrix} 1-\lambda & -N^* a \\ 1-\dfrac{1}{r} & \dfrac{N^* a}{r} - \lambda \end{vmatrix} = \lambda^2 - \left(1 + \frac{N^* a}{r}\right)\lambda + N^* a = 0 \tag{6.96}$$

En utilisant la valeur de N^* donnée dans (6.87), on montre que les deux zéros sont complexes conjugués et que leur produit, $N^* a$, qui est aussi égal au carré de leur module commun, est strictement supérieur à 1 pour $r > 1$. On en déduit que cet équilibre est toujours instable. Il est donc illusoire de vouloir représenter un système réel présentant un équilibre stable avec ce modèle. De plus, les simulations numériques montrent que l'équilibre non trivial (N^*, P^*) est également instable pour des perturbations d'amplitude finie ; la solution croît sans borne. Le modèle ne peut donc être appliqué tel quel à aucun système réel.

Une modification possible de de (6.86) consiste à introduire une saturation dans la dynamique des proies, par exemple, en y incorporant la loi de Ricker. On aura alors

$$\begin{cases} N_{k+1} = N_k \exp\left[r\left(1 - \dfrac{N_k}{K}\right) - a\, P_K\right] \\ \\ P_{k+1} = N_k \left[1 - \exp(-a\, P_K\right] \end{cases} \tag{6.97}$$

dont la dynamique peut présenter des configurations d'équilibre stable.

Chapitre 7

Modélisation au moyen d'équations aux dérivées partielles.

7.1 Dynamique de population avec distribution d'âge.

Le modèle de croissance logistique avec retard (5.22) peut être généralisé pour tenir compte du fait que la dynamique de la population à un instant t donné dépend d'une certaine moyenne des populations aux instants précédents. On aura alors l'équation intégro-différentielle

$$\frac{dN}{dt} = r\,N(t)\left[1 - \frac{1}{K}\int_0^\infty W(\tau)N(t-\tau)d\tau\right] \tag{7.1}$$

où $W(\tau)$ désigne une fonction de pondération décrivant l'influence de la taille des populations aux instants passés sur la disponibilité actuelle ou sur la qualité des ressources. L'équation (5.22) constitue un cas particulier de (7.1) où la fonction $W(\tau)$ est non nulle pour le seul retard T. De même, l'équation (5.5) s'obtient à partir de (7.1) en posant $W(\tau) = \delta(\tau)$.

De façon alternative, on peut interpréter le terme entre crochets dans (7.1) comme un terme correctif au terme Malthusien rN tenant compte de la distribution des âges au sein de la population ; une population trop jeune, immature ou au contraire trop âgée présentant des taux de reproduction et de croissance faibles. Dans cette optique, on peut rendre plus explicite la dépendance de la dynamique en la structure de la population en décrivant explicitement sa distribution d'age.

Soit $n(a,t)$ la densité de population au temps t en fonction de l'âge a, *i.e.* $n(a,t)$ est telle que le nombre d'individus dont les âges sont compris entre a_1 et $a_2 (> 0)$ est donné par

$$\int_{a_1}^{a_2} n(a,t)da \tag{7.2}$$

Soit $b(a)$ et $m(a)$ respectivement les taux de reproduction et de mortalité en fonction de l'âge. Les individus d'âge a au temps t forment la classe d'âge $a+\Delta t$ au temps $t+\Delta t$

mais leur nombre est réduit par la mortalité. On a donc

$$n(a+\Delta t,t+\Delta t)-n(a,t)=-\int_0^{\Delta t} m(a+\tau)n(a+\tau,t+\tau)d\tau \tag{7.3}$$

Soit encore

$$\left(\frac{\partial n}{\partial a}(a,t)+\frac{\partial n}{\partial t}(a,t)\right)\Delta t+o(\Delta t^2)=-m(a)n(a,t)\Delta t+o(\Delta t^2) \tag{7.4}$$

En passant à la limite pour $\Delta t \to 0$, il vient

$$\frac{\partial n}{\partial a}+\frac{\partial n}{\partial t}=-m(a)n \tag{7.5}$$

Le nombre de naissances est donné par

$$n(0,t)=\int_0^{a_{\max}} b(a)n(a,t)da \tag{7.6}$$

où $a_{\max}$ désigne l'âge maximum pour lequel $b(a)$ est non nul.

Les équations (7.5) et (7.6) constituent le modèle d'évolution de la population. Celui-ci est donc entièrement décrit par la donnée des lois $b(a)$ et $m(a)$. En plus de (7.6), il convient d'imposer une condition initiale du type

$$n(a,0)=f(a) \tag{7.7}$$

donnant la distribution des âges à l'instant initial.

7.1.1 Solution générale.

Le problème (7.5)-(7.7) peut être résolu en remarquant que, dans l'espace (a,t), un même individu évolue le long d'une droite $a-t=C^{\text{te}}$. Le long de cette courbe 'caractéristique', l'équation (7.5) s'écrit

$$\frac{d}{dt}n(a(t),t)=-m(a)n(a(t),t) \tag{7.8}$$

dont la solution est simplement de la forme

$$n(a,t)\propto \exp\left[-\int_a^{a+t} m(s)ds\right] \tag{7.9}$$

L'espace (a,t) est divisé en deux par la caractéristique $a=t$ (Fig. 7.1).

I.) Pour $a>t$, la solution est entièrement déterminée par la condition initiale (7.7), *i.e.*

$$n(a,t)=f(a-t)\exp\left[-\int_{a-t}^{a} m(s)ds\right] \quad (a>t) \tag{7.10}$$

Cette partie de la solution décrit le vieillissement et la mortalité progressive de la population initiale.

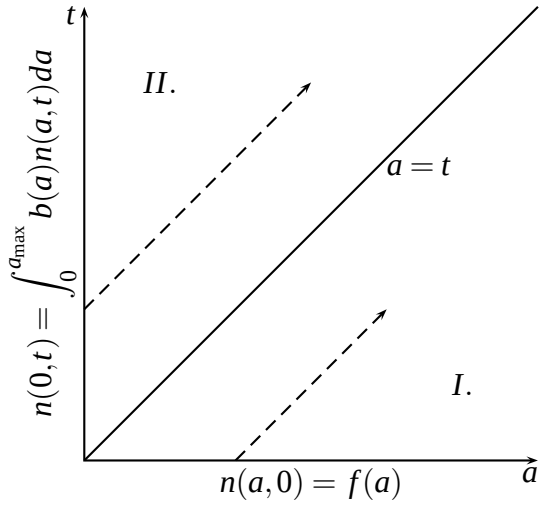


FIG. 7.1

II.) Pour $a < t$, par contre, la condition initiale (7.7) est sans influence ; les individus concernés sont tous nés entre l'instant initial $t = 0$ et l'instant considéré. La solution est donc conditionnée par la condition auxiliaire (7.6) qui sert de condition 'initiale' de sorte que

$$n(a,t) = n(0,t-a)\exp\left[-\int_0^a m(s)ds\right] \quad (a < t) \tag{7.11}$$

C'est cette dernière équation qui conditionne le comportement de la solution $n(a,t)$ à long terme.

7.1.2 Solution auto-similaire.

Examinons la possibilité d'occurrence d'une solution du type

$$n(a,t) = e^{\gamma t}g(a) \tag{7.12}$$

Une telle solution auto-similaire est caractérisée par une structure des âges constante à tous les instants ; seul le nombre total d'individus varie au cours du temps.

Substituant (7.12) dans (7.6) et (7.11), on a

$$g(a) = g(0)\exp\left[-\gamma a - \int_0^a m(s)ds\right] \tag{7.13}$$

et

$$\begin{aligned} g(0) &= \int_0^{a_{\max}} b(a)g(a)da \\ &= g(0)\int_0^{a_{\max}} b(a)\exp\left[-\gamma a - \int_0^a m(s)ds\right]da \end{aligned} \tag{7.14}$$

Posant

$$\phi(\gamma) = \int_0^{+a_{\max}} b(a) \exp\left[-\gamma a - \int_0^a m(s)ds\right] da \tag{7.15}$$

on observe que (27) ne représente une solution du problème que si γ est la solution unique de l'équation[1]

$$\phi(\gamma) = 1 \tag{7.16}$$

Si $\phi(0) > 1$, la solution de (7.16) définit une valeur $\gamma > 0$ et la population croît exponentiellement au cours du temps. Si $\phi(0) < 1$, la solution de (7.16) correspond à une valeur négative de γ et la population décline exponentiellement.

Considérons le cas particulier où $b(a)$ est non nul uniquement au voisinage de $a = a_0$ et où la mortalité est une fonction linéaire de l'âge, *i.e.* $b(a) = \mu\delta(a - a_0)$ et $m(a) = m_0 + \alpha a$. Il vient

$$\begin{aligned}
\phi(\gamma) &= \int_0^{+\infty} \mu\delta(a-a_0) \exp\left[-\gamma a - \int_0^a m(s)ds\right] da \\
&= \int_0^{+\infty} \mu\delta(a-a_0) \exp\left[-\gamma a - m_0 a - \alpha\frac{a^2}{2}\right] da \\
&= \mu \exp\left[-\gamma a_0 - m_0 a_0 - \alpha\frac{a_0^2}{2}\right]
\end{aligned} \tag{7.17}$$

La population ne pourra donc se maintenir que si

$$\mu \exp\left[-m_0 a_0 - \alpha\frac{a_0^2}{2}\right] \geq 1 \tag{7.18}$$

7.2 Advection unidimensionnelle.

L'advection unidimensionnelle d'un constituant passif, *i.e.* ne subissant aucune transformation physique, biologique ou chimique, donne lieu à une équation semblable à (7.5). Si $C(x,t)$ désigne la concentration au point x au temps t, on a, en effet[2],

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = 0 \tag{7.19}$$

où $u(> 0)$ désigne la vitesse du courant responsable de l'advection du constituant considéré. Pour résoudre cette équation, on doit logiquement disposer de conditions initiales

$$C(x,0) = f(x) \tag{7.20}$$

décrivant la distribution de la concentration au temps $t = 0$ et d'une condition limite

$$C(0,t) = g(t) \tag{7.21}$$

[1] Cette équation admet une solution unique puisque $\phi(\gamma)$ est une fonction décroissante de γ.

[2] Cette équation sera établie plus loin dans le cas général tridimensionnel.

fixant la valeur de la concentration à la limite amont (prise ici en $x = 0$) du domaine spatial considéré.

Ici encore, la solution est déterminée par les conditions initiales et les conditions aux limites du problème transportées le long des courbes $x - ut = C^{te}$ (Fig. 7.2) :

$$C(x,t) = C(x - u(t-\tau), t-\tau), \qquad \forall\tau \tag{7.22}$$

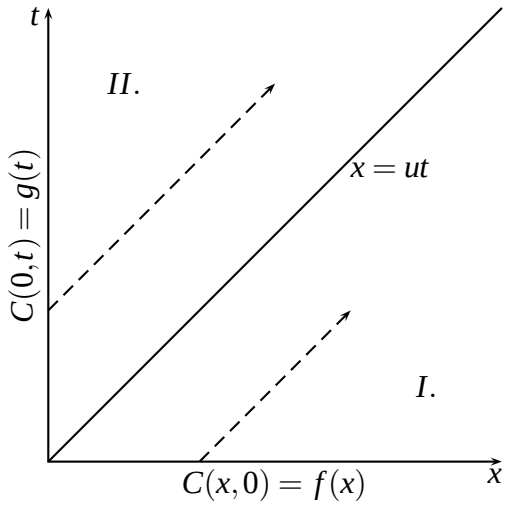


FIG. 7.2

Dans le domaine I, situé sous la droite $x = ut$, la solution est entièrement déterminée par la distribution initiale $f(x)$, laquelle est simplement translatée d'une distance ut, *i.e.*

$$C(x,t) = f(x - ut), \qquad \text{pour } x > ut \tag{7.23}$$

Dans le domaine II, *i.e.* pour des temps supérieurs à x/u, la solution ne dépend que de la condition à la frontière amont, *i.e.*

$$C(x,t) = g(t - x/u), \qquad \text{pour } x < ut \tag{7.24}$$

7.3 Généralisation et classification des EDP.

Les droites $a - t = C^{te}$ et $x - ut = C^{te}$ apparaissant dans les deux premières sections sont connues sous le nom de 'caractéristiques'. Dans le cas général, les caractéristiques d'une EDP ou d'un système d'EDP sont des courbes ou des surfaces de l'espace des variables indépendantes qui admettent la double interprétation suivante.

i. Les caractéristiques sont des courbes (ou des surfaces) le long desquelles se propagent naturellement l'information et, en particulier, les conditions initiales.

ii. Il est impossible d'extrapoler la solution au travers des caractéristiques.

Ces propriétés complémentaires se retrouvent bien dans la discussion du modèle avec structure d'âge et dans le modèle d'advection unidimensionnelle.

L'existence de caractéristiques permet de classer les EDP. Considérons l'équation du second ordre

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + e\frac{\partial u}{\partial x} + f\frac{\partial u}{\partial y} + g\,u = 0 \tag{7.25}$$

où a, b, c, d, e, f et g sont des fonctions réelles des variables indépendantes x et y.

- Si $ac < b^2$, l'équation possède deux familles de caractéristiques réelles. Les deux caractéristiques qui se croisent en un point donné sont chacune porteuse d'une partie de la solution. La combinaison de ces deux éléments détermine complètement la solution. L'équation différentielle est dite hyperbolique.
  L'équation hyperbolique du second ordre la plus simple est

$$\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = 0 \tag{7.26}$$

  Celle-ci décrit la propagation d'ondes à la vitesse c dans un milieu unidimensionnel.
- Si $ac < b^2$, l'équation ne possède pas de caractéristiques (ou plus exactement, les deux familles de caractéristiques sont complexes). L'équation est dite elliptique.
  Le modèle de ce type d'équation est l'équation de Laplace

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{7.27}$$

  Celle-ci décrit les déformations d'une membrane tendue sur une courbe de support.
- Si $ac = b^2$, l'équation admet une seule famille de caractéristiques réelles. Elle est dite parabolique.
  Il en est par exemple ainsi de l'équation de la chaleur

$$\frac{\partial T}{\partial t} = k\frac{\partial^2 T}{\partial x^2} \tag{7.28}$$

  décrivant la diffusion de la chaleur dans un milieu unidimensionnel au repos.

Il est à noter que le type d'une EDP est entièrement déterminé par les coefficients des dérivées d'ordre les plus élevés. Comme ceux-ci peuvent varier d'un point à l'autre, le type d'une EDP peut également varier d'un point à l'autre.

Les solutions des équations hyperboliques, elliptiques et paraboliques présentent des comportements différents.

7.3.1 Conditions initiales ou aux limites.

Les conditions initiales et aux limites sont généralement d'un des trois types suivants :

- *Condition de Dirichlet* : la valeur du champ inconnu u est imposée sur un segment de la frontière du domaine étudié.
- *Condition de Neumann* : la dérivée du champ selon la normale à la frontière est imposée.
- *Condition de Robin (ou de Newton)* : on impose une combinaison des valeurs de u et de sa dérivée.

Le nombre total de conditions initiales ou aux limites nécessaires pour définir une solution unique d'un problème aux dérivées partielles est égal (sauf exception) à la somme des ordres de dérivation maximaux par rapport à chacune des variables.

En général, pour chaque variable indépendante, on fixe un nombre de conditions initiales/aux limites égal à l'ordre maximum de dérivations par rapport à cette variable. Différentes combinaisons de conditions initiales/aux limites sont cependant généralement appliquées aux équations des trois types.

Problème hyperbolique.

Examinons d'abord l'équation hyperbolique

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial u}{\partial x^2} = 0 \tag{7.29}$$

et sa discrétisation naturelle

$$\frac{u_i^{k+1} - 2u_i^k + u_i^{k-1}}{\Delta t^2} - c^2 \frac{u_{i+1}^k - 2u_i^k + u_{i-1}^k}{\Delta x^2} = 0 \tag{7.30}$$

si on note

$$u_i^k = u(t_0 + k\Delta t, x_0 + i\Delta x) \tag{7.31}$$

Cette discrétisation suggère de faire avancer la solution dans le temps selon la récurrence

$$u_i^{k+1} = f(u_i^{k-1}, u_{i-1}^k, u_i^k, u_{i+1}^k) \tag{7.32}$$

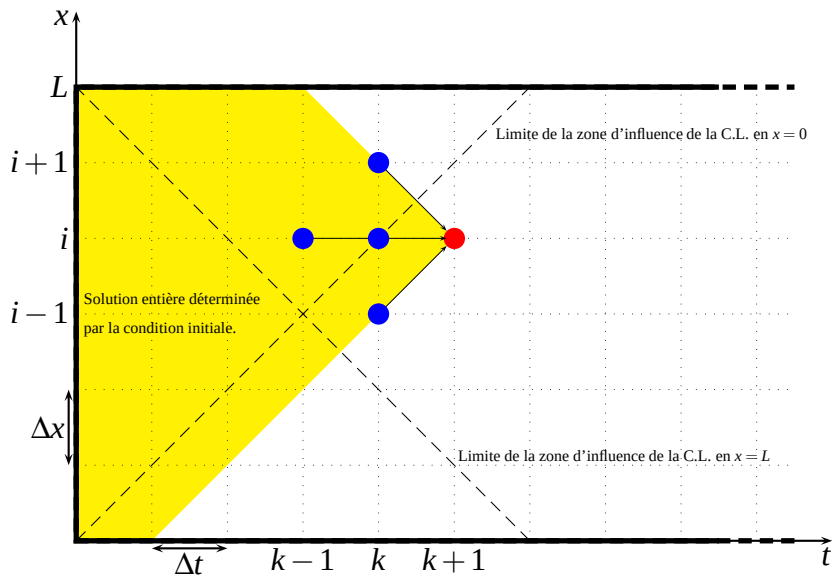


FIG. 7.3 – Discrétisation de l'équation des ondes (7.29) dans un domaine spatial fini.

La figure 7.3 illustre la dépendance de u_i^{k+1} en les valeurs discrètes aux instants précédents dans le cas d'un problème résolu dans un domaine spatial $x \in [0,L]$ de dimension finie. Les valeurs de l'inconnue u_i^{k+1} au nouveau pas de temps dépendent des valeurs déjà calculées appartenant au cône de dépendance représenté en jaune sur la figure. Pour que la discrétisation soit stable, il faut que les caractéristiques réelles ($x \pm ct = C^{\text{te}}$) appartiennent au cône de dépendance de la récurrence.

On constate que le calcul de la solution pour $k = 1$ demande de disposer des valeurs de l'inconnues en $k = 0$ et $k = -1$. Ceci n'est possible que si deux conditions auxiliaires fournissent ces valeurs. Deux conditions initiales doivent donc être fixées en $t = 0$, par exemple en imposant la distribution initiale de u et de sa dérivée temporelle , *i.e.*

$$u(0,x) = f(x); \frac{\partial u}{\partial t}(0,x) = g(x) \tag{7.33}$$

De même, si l'indice $i = 0$ correspond à une frontière du domaine spatial, le calcul de la solution au voisinage de cette frontière (*i.e.* $i = 1$) demande qu'une information soit disponible sur la frontière. Une condition aux limites doit être donnée en chaque point de la frontière spatiale du domaine d'intégration, *i.e.* une condition en $x = 0$ et une autre en $x = L$. Par exemple, on imposera

$$u(t,0) = h_1(t), \qquad u(t,L) = h_2(t) \tag{7.34}$$

où $h_1(t)$ et $h_2(t)$ sont des fonctions connues du temps.

Remarquons que le nombre de conditions aux limites nécessaires à la résolution du problème est bien égal à l'ordre des dérivées partielles de chaque type apparaissant dans (7.29).

Sur la figure 7.3, on a aussi représenté les limites des zones d'influence des conditions aux limites du problème. Ces zones sont limitées par des caractéristiques du problème différentiel. On remarque que la condition initiale détermine complètement la solution dans un domaine triangulaire adjacent à l'axe $t = 0$. Dans le cas d'un domaine spatial infini, ce triangle couvre l'ensemble de l'espace et la solution ne dépend que des conditions initiales.

Problème elliptique.

Le problème elliptique

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{7.35}$$

donne lieu à la discrétisation

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2} + \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2} = 0 \tag{7.36}$$

où

$$u_{i,j} = u(x_0 + i\Delta x, y_0 + j\Delta j) \tag{7.37}$$

soit

$$u_{i,j} = f(u_{i+1,j}, u_{i-1,j}, u_{i+1,j}, u_{i,j-1}) \tag{7.38}$$

La figure 7.4 illustre la dépendance entre les variables résultant de cette discrétisation.

La valeur de l'inconnue en un point est une fonction de toutes les valeurs voisines. Il n'y a pas de direction préférentielle de propagation de l'information (pas de caractéristique).

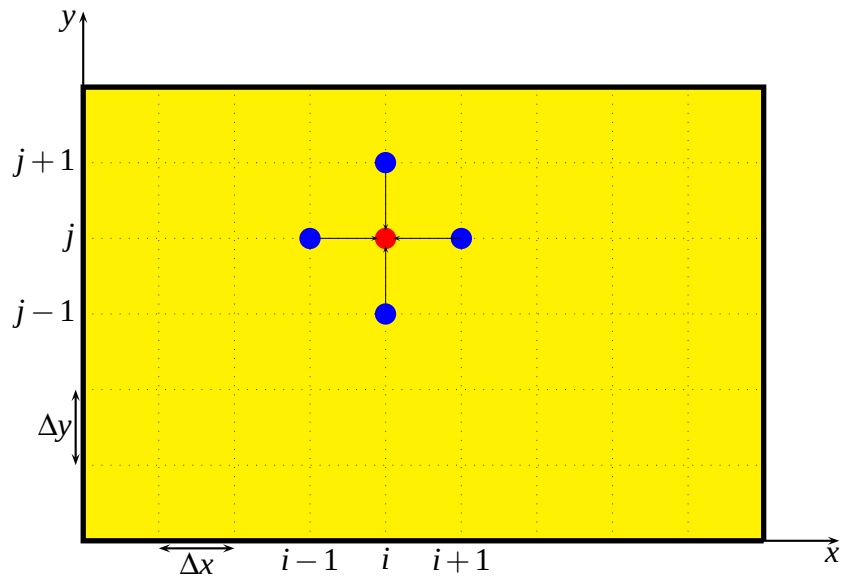


FIG. 7.4 – Discrétisation de l'équation elliptique (7.35) dans un domaine fini.

Une solution unique peut être générée à l'intérieur d'un domaine borné en tout point de la frontière duquel une condition limite unique est imposée.

Problème parabolique.

L'équation parabolique

$$\frac{\partial u}{\partial t} = \kappa \frac{\partial^2 u}{\partial x^2} \tag{7.39}$$

donne lieu à la discrétisation

$$\frac{u_i^{k+1} - u_i^k}{\Delta t} = \kappa \frac{u_{i+1}^k - 2u_i^k + u_{i-1}^k}{\Delta x^2} \tag{7.40}$$

Ceci conduit à une dépendance semblable à celle de l'équation hyperbolique, soit

$$u_i^{k+1} = f(u_i^k, u_{i+1}^k, u_{i-1}^k) \tag{7.41}$$

Pour des raisons de stabilité de schéma numérique, il est cependant généralement nécessaire de traiter les équations paraboliques de façon implicite (au moins partiellement). En effet, on associe une vitesse de propagation infinie à l'équation (7.39) : toute perturbation en un point se fait sentir instantanément en tous les points du domaine. Afin de pouvoir représenter cet effet numériquement, on écrira

$$\frac{u_i^{k+1} - u_i^k}{\Delta t} = \frac{\kappa}{\Delta x^2}\left[\alpha\left(u_{i+1}^k - 2u_i^k + u_{i-1}^k\right) + (1-\alpha)\left(u_{i+1}^{k+1} - 2u_i^{k+1} + u_{i-1}^{k+1}\right)\right] \tag{7.42}$$

Dès lors, toutes les variables d'un même pas de temps sont liées entre-elles par des équations algébriques du type

$$g(u_{i-1}^{k+1}, u_i^{k+1}, u_{i+1}^{k+1}) = f(u_i^k, u_{i+1}^k, u_{i-1}^k) \tag{7.43}$$

En raison de ces équations, toutes les valeurs au nouveau pas de temps $k+1$ doivent être déterminées simultanément.

La dépendance entre les valeurs de la discrétisation est illustrée à la figure 7.5.

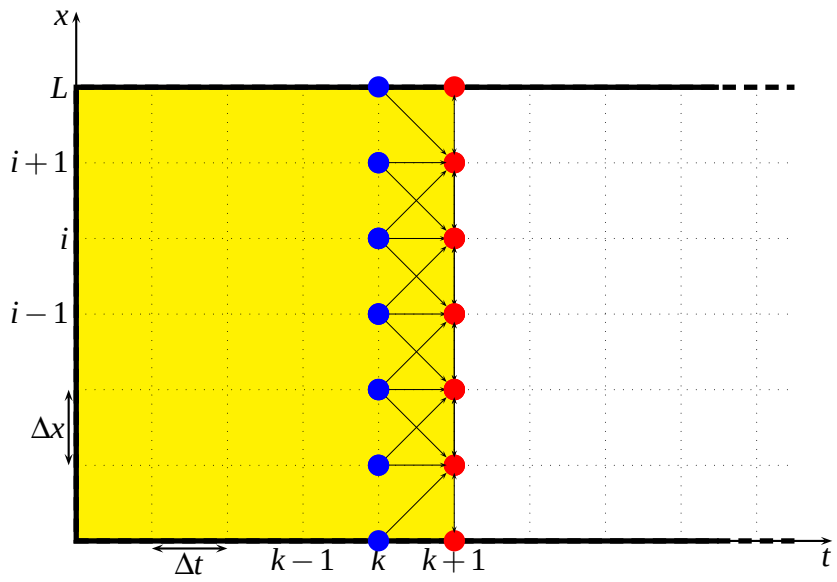


FIG. 7.5 – Discrétisation partiellement implicite de l'équation de la chaleur (7.39) dans un domaine spatial fini.

Cette fois, une seule condition initiale doit être imposée. Dans un domaine borné ($x \in [0,L]$), une condition limite supplémentaire doit être appliquée en chacune des extrémités de l'intervalle $[0,L]$, ou, plus généralement, en chaque point de la frontière. Ceci permet de déterminer complètement la solution pour tout $t > 0$.

7.4 Modèle 1D d'advection-diffusion-migration.

Le mouvement des substances dissoutes ou en suspension ainsi que des organismes présents dans le milieu marin peut être représenté comme la combinaison de trois processus.

– On désigne par *advection* le mouvement qui résulte du transport ordonné par le fluide en mouvement ; celui-ci est donc caractérisé par la vitesse w du fluide.

– La *diffusion* représente l'effet résultant des mouvements aléatoires des différentes particules étudiées qui se superpose au mouvement d'ensemble (advection). La diffusion ne s'accompagne d'aucun mouvement net mais entraîne, généralement, une homogénéisation des propriétés ; elle est caractérisée par un coefficient de diffusion λ.
– La *migration* et la *sédimentation* représentent les mouvements ordonnés, indépendants du mouvement d'ensemble du fluide. Certaines espèces sont ainsi capables de mouvement propres à la recherche de nourriture ou de conditions environnementales favorables. De même, les particules plus denses (ou plus légères) que l'eau dans laquelle elles baignent subissent des mouvements verticaux propres. Ces processus sont caractérisés par une vitesse de migration ou de sédimentation w_s.

Si on s'intéresse aux seuls mouvements verticaux dans la colonne d'eau, l'équation différentielle décrivant ces différents processus s'écrit[3], pour la concentration C d'une grandeur quelconque,

$$\frac{\partial C}{\partial t} + (w + w_s)\frac{\partial C}{\partial z} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial C}{\partial z}\right) \tag{7.44}$$

où t désigne le temps et z représente la coordonnée verticale (croissante vers le haut).

En l'absence de terme de diffusion ($\lambda = 0$), l'équation 7.44 se réduit à

$$\frac{\partial C}{\partial t} + (w + w_s)\frac{\partial C}{\partial z} = 0 \tag{7.45}$$

qui est en tout point semblable à (7.5) avec une mortalité nulle. L'équation est donc de nature hyperbolique et décrit le transport passif le long de la caractéristique

$$z - (w + w_s)t = \text{constante} \tag{7.46}$$

En d'autres termes, la distribution initiale est simplement translatée à la vitesse $(w + w_s)$.

En l'absence de terme d'advection et de migration, ($w + w_s = 0$), l'équation (7.44) se réduit à

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z}\left(\lambda \frac{\partial C}{\partial z}\right) \tag{7.47}$$

Elle est donc de nature parabolique et décrit la distribution progressive de la distribution initiale.

Le modèle 1D vertical (7.44) s'appliquant généralement à la colonne d'eau, il doit être résolu dans un domaine limité par la surface et le fond. Différentes conditions limites peuvent alors être appliquées selon la nature de la grandeur C étudiée. Remarquons cependant que l'équation est du second ordre par rapport à la coordonnée spatiale et requiert donc deux conditions aux limites (en chaque instant) en plus de la donnée de la distribution initiale en tout point de la colonne d'eau.

[3]La dérivation de cette équation sera explicitée plus loin dans le cas général tridimensionnel.

Considérons tout d'abord le cas de la température $C = T$. Dans ce cas, on a évidemment $w_s = 0$. Au fond, on considère généralement que le flux de chaleur est nul. Dès lors on imposera la condition limite

$$wT - \lambda \frac{\partial T}{\partial z}\Big|_{fond} = 0 \tag{7.48}$$

correspondant à l'annulation du flux total (advectif + diffusif). Par continuité, la vitesse verticale du fluide doit elle-même s'annuler au fond (supposé horizontal). Dès lors, cette condition se réduit à

$$-\lambda \frac{\partial T}{\partial z}\Big|_{fond} = 0 \tag{7.49}$$

La même condition peut être appliquée en surface (tenant compte également de l'annulation de la vitesse verticale à cet endroit). Cependant, pour tenir compte de l'échange de chaleur avec l'atmosphère, on écrira généralement

$$-\lambda \frac{\partial T}{\partial z}\Big|_{surface} = \frac{1}{\rho c_p} J_{heat} \tag{7.50}$$

où J_{heat} désigne le flux de chaleur en surface (quantité de chaleur traversant la surface par unité de surface horizontale et par unité de temps), ρ et c_p représentent la masse par unité de volume et la chaleur massique de l'eau. Ce flux de chaleur dépend à la fois de l'état de l'atmosphère et de la température de l'eau en surface. En effet, l'échange de chaleur sensible entre la colonne d'eau et l'air dépend de la différence de température entre ces deux milieux. De même, l'évaporation et les radiations en ondes longues issues des couches superficielles dépend de la température de l'eau et de l'humidité spécifique des l'air environnant. La prise en compte des différents facteurs influençant l'échange thermique est extrêmement complexe. Une modélisation simple de ces échanges est cependant possible sous la forme

$$-\lambda \frac{\partial T}{\partial z}\Big|_{surface} = Q + \alpha(T_{air} - T_{surface}) \tag{7.51}$$

où les coefficients Q et α permettent une modélisation globale des différents processus responsables de l'échange.

Dans le cas où les données sont insuffisantes pour évaluer le flux de chaleur en surface (ou les coefficients Q et α de la formulation (7.51)) mais que les observations satellitaires décrivent l'évolution de la température en surface, on peut également écrire

$$-\lambda \frac{\partial T}{\partial z}\Big|_{surface} = \alpha(T_{obs} - T_{surface}) \tag{7.52}$$

qui introduit un terme de rappel ('nudging') de la température calculée $T_{surface}$ vers la température observée T_{obs}. L'intensité de ce terme de rappel est gouvernée par le paramètre α représentant l'inverse du temps caractéristique du rappel.

Dans le cas de la concentration de particules soumises à la sédimentation ($w_s < 0$), plusieurs possibilités peuvent apparaître. Ainsi, si les conditions hydrodynamiques ne permettent pas le dépôt par sédimentation sur le fond, on aura (tenant compte de $w = 0$ au fond)

$$w_s C - \lambda \frac{\partial C}{\partial z}\Big|_{fond} = 0 \tag{7.53}$$

Si les conditions sont calmes et que les particules se déposent sur le fond, on écrira par contre

$$\lambda \frac{\partial C}{\partial z}\Big|_{fond} = 0 \tag{7.54}$$

de sorte que le flux de dépôt sur le fond est donné par $w_s C$. Si un flux extérieur $J_{surface}$ est imposé en surface, par exemple lié aux particules se déposant sur la surface en provenance de l'atmosphère, on aura

$$w_s C - \lambda \frac{\partial C}{\partial z}\Big|_{surface} = J_{surface} \tag{7.55}$$

7.5 Modèle général tridimensionnel.

L'équation générale gouvernant la dynamique de toutes les substances présentes dans le milieu marin peut être établie à partir d'un bilan de masse de cette substance sur un volume de référence quelconque.

7.5.1 Équation de continuité.

Avant de considérer une substance quelconque, examinons d'abord le cas particulier de l'eau elle-même. Délimitons, par la pensée, un volume de référence V, fixe mais de forme quelconque. Soit Σ la surface latérale de ce volume et $\mathbf{n}$ la normale unitaire extérieure. La masse totale du fluide comprise dans le volume de contrôle est donnée par

$$\mathcal{M}(t) = \iiint_{\Omega} \rho dV \tag{7.56}$$

où ρ désigne la masse par unité de volume. Le transport du fluide au travers de la surface latérale Σ du volume de contrôle constitue la seule cause possible de variation de $\mathcal{M}(t)$. Pendant un petit instant dt, la masse de fluide traversant un petit élément $d\Sigma$ de normale $\mathbf{n}$, est donnée par

$$\rho \mathbf{n} \cdot \mathbf{v} d\Sigma dt \tag{7.57}$$

où $\mathbf{v}$ et ρ désignent la vitesse et la masse du fluide à l'endroit et à l'instant considérés. En travaillant par unité de temps et en intégrant sur la surface latérale totale du volume de contrôle, on obtient

$$\frac{d}{dt}\mathcal{M}(t) = -\iint_{\Sigma} \rho \mathbf{v} \cdot \mathbf{n} \; d\Sigma \tag{7.58}$$

Remarquons que le signe moins devant l'intégrale du membre de droite trouve ici sa justification puisque l'augmentation de la masse $\mathcal{M}$ s'accompagne d'un flux net négatif au travers de la surface Σ. Il vient donc

$$\frac{d}{dt}\iiint_{\Omega}\rho dV = -\iint_{\Sigma}\rho\mathbf{v}\cdot\mathbf{n}\ d\Sigma \tag{7.59}$$

qui constitue l'expression intégrale de la *loi de conservation de la masse* ou *loi de continuité.*

Une expression local de la continuité peut être obtenue en utilisant le théorème de Gauss, lequel établit (sous certaines conditions qui seront considérées rencontrées ici) l'égalité

$$\iiint_{V}\nabla\cdot\mathbf{F}dV = \iint_{\Sigma}\mathbf{F}\cdot\mathbf{n}d\Sigma \tag{7.60}$$

pour tout champ vectoriel $\mathbf{F}$ et tout volume V de surface Σ et de normale extérieure $\mathbf{n}$.

Appliquant ce résultat pour transformer le membre de droite de (7.59), on obtient

$$\iiint_{\Omega}\left(\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v})\right)\ dV = 0 \tag{7.61}$$

En dérivant la relation (7.61), aucune hypothèse n'a été faite sur le volume de contrôle. Celui-ci est quelconque. Aussi, l'annulation de l'intégrale dans (7.61) ne peut résulter d'un choix particulier du volume de contrôle mais implique l'annulation de l'intégrand lui-même. Dès lors, on obtient la forme locale de l'équation de continuité

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v}) = 0 \tag{7.62}$$

En développant le second terme, l'équation de continuité peut encore s'écrire

$$\frac{\partial\rho}{\partial t}+\mathbf{v}\cdot\nabla\rho+\rho\nabla\cdot\mathbf{v} = 0 \tag{7.63}$$

ou encore

$$D_t\rho+\rho\nabla\cdot\mathbf{v} = 0 \tag{7.64}$$

en introduisant l'opérateur de dérivée matérielle (ou totale)

$$D_t(\ldots) = \frac{\partial}{\partial t}(\ldots)+\mathbf{v}\cdot\nabla(\ldots) \tag{7.65}$$

qui mesure le taux de variation d'une grandeur pour une particule fluide donnée.

La forme (7.64) est bien adaptée pour faire apparaître les simplifications appropriées à l'océanographie. En effet, en bonne approximation, l'eau de mer, mélange d'un grand nombre de substances différentes, peut être considérée incompressible[4] de sorte que $D_t\rho = 0$ et donc

$$\nabla\cdot\mathbf{v} = 0 \tag{7.66}$$

[4]Dans l'étude de l'hydrodynamique marine, on ignore les variations de densité de l'eau de mer sauf dans la projection de l'équation de quantité de mouvement sur la verticale où les petites variations de densité sont amplifiées par la multiplication par l'accélération de pesanteur g. Ceci donne naissance au concept de poussée.

7.5.2 Équation de bilan.

Le bilan de masse des substances (ou organismes) dissoutes ou en suspension peut être formalisé comme dans la section précédente. Soit C la concentration volumique, *i.e.* par unité de volume, de la substance étudiée.

La masse totale de cette substance dans un volume de contrôle quelconque est donnée par

$$\mathcal{M}_C(t) = \iiint_\Omega C dV \tag{7.67}$$

Évaluons maintenant les taux de variations de cette masse induits par les différents processus en jeu.

- La substance est d'abord transportée par le fluide au travers de la frontière Σ du volume de contrôle. Le flux total (masse par unité de temps) correspondant est donné par l'intégrale
$$-\iint_\Sigma C\mathbf{v}\cdot\mathbf{n}\, d\Sigma \tag{7.68}$$
où le signe négatif est introduit pour qu'une valeur positive de ce terme corresponde à une augmentation de la masse au sein du volume de contrôle.
- La substance est également soumise à la diffusion. Celle-ci représente les mouvements désordonnés des particules fluides qui ne produisent aucun mouvement net mais induisent, à petite échelle du moins, l'homogénéisation progressive des propriétés du fluide. Selon le modèle classique de Fourier-Fick, le flux de diffusion $\mathbf{J}_{diff}$ est supposé proportionnel au gradient de la propriété étudiée, soit
$$\mathbf{J}_C^{diff} = -\lambda\nabla C \tag{7.69}$$
où le coefficient λ est le *coefficient de diffusion*. Le signe négatif est introduit dans cette expression car le transport s'effectue de la zone de plus forte concentration vers les zones où la concentration est plus faible et s'effectue donc dans le sens opposé à celui du gradient (pour rappel, le gradient pointe toujours dans la direction de croissance la plus rapide de la grandeur).
Dans le cas de l'équation de continuité, ce terme était absent puisque la diffusion de l'eau dans l'eau ne correspond à aucun transfert de masse.
Intégrant les flux de diffusion sur la surface totale du volume de contrôle, il vient
$$-\iint_\Sigma \mathbf{J}_C^{diff}\cdot\mathbf{n}\, d\Sigma = \iint_\Sigma \lambda\nabla C\cdot\mathbf{n}\, d\Sigma \tag{7.70}$$
- Les termes de sédimentation et migration éventuels se traitent de la même façon que le terme d'advection en remplaçant la vitesse $\mathbf{v}$ du fluide par la vitesse de migration/sédimentation $\mathbf{v}_{s/m}$.
- La substance peut également subir des transformation physico-chimiques avec ou sans interaction avec d'autres substances. De même, les organismes peuvent interagir entre-eux. La modélisation et l'étude de ces interactions à fait l'objet du chapitre 6. Ici, nous représenterons leur effet sous la forme d'un terme de

production-destruction dont le taux par unité de volume est donné par Q_c. Ce taux est positif dans le cas d'une production et négatif dans le cas de la disparition du constituant C. Dans le volume de contrôle, on a donc

$$\iiint_V Q_c dV \tag{7.71}$$

En regroupant tous les termes introduits ci-dessus, le bilan global de la substance C dans le volume de contrôle s'écrit

$$\frac{d}{dt}\iiint_\Omega C dV = -\iint_\Sigma C(\mathbf{v}+\mathbf{v}_{s/m})\cdot\mathbf{n}\, d\Sigma + \iint_\Sigma \lambda\nabla C\cdot\mathbf{n}\, d\Sigma + \iiint_V Q_c dV \tag{7.72}$$

Cette expression intégrale peut être transformée en utilisant le théorème de Gauss (7.60),

$$\iiint_\Omega \left[\frac{\partial C}{\partial t} + \nabla\cdot\big(C(\mathbf{v}+\mathbf{v}_{s/m})\big) - \nabla\cdot\big(\lambda\nabla C\big) - Q_c\right] dV = 0 \tag{7.73}$$

Le volume de contrôle étant arbitraire, on en déduit, comme précédemment,

$$\frac{\partial C}{\partial t} + \nabla\cdot\big(C(\mathbf{v}+\mathbf{v}_{s/m})\big) - \nabla\cdot\big(\lambda\nabla C\big) - Q_c = 0 \tag{7.74}$$

qui représente l'expression générale de l'équation de bilan d'une substance quelconque dans le milieu marin.

L'équation (7.74) peut être par particularisée pour les différentes substances en adaptant (et en introduisant au besoin une paramétrisation adaptée) les expressions du taux de production-destruction, de la vitesse de sédimentation (et éventuellement du coefficient de diffusion).

Les substances pour lesquelles $Q_c = 0$ sont qualifiées de *passives*. Celles-ci sont simplement transportées par le fluide sans subir aucune transformation.

Les substances dont la dynamique ne dépend que de leur propre concentration sont dites *semi-passives*. Ainsi en est-il des substances qui subissent une désintégration radioactive ($Q_c = -\alpha C$). Leur évolution peut être simulée indépendamment de celle des autres substances.

Enfin, les substances *actives* interagissent entre-elles et leur dynamique ne peut donc être appréhendée que globalement.

7.5.3 Intégration dans l'espace d'état.

L'équation (7.74) est valable pour n'importe qu'elle substance. Or, l'eau de mer est un mélange d'un ensemble incommensurable de substances différentes qu'il est impossible de décrire simultanément dans un seul modèle. Pour obtenir un modèle mathématique réellement utile et aisément manipulable, il est essentiel de conserver uniquement un nombre réduit de variables d'état, *i.e.* de réduire la *portée* de ce modèle. C'est l'essence même de la démarche de modélisation que de sélectionner un ensemble limité de variables d'état permettant de décrire le comportement du système réel de façon adéquate tout en

limitant la taille du modèle pour en permettre la calibration, la validation, la résolution numérique et l'interprétation des résultats.

La complexité d'un modèle dépend évidemment de la complexité, et en particulier de la non linéarité, de la dynamique du système réel. Elle dépend également du propos et de l'objectif de l'étude réalisée. S'agit-il d'une étude scientifique, d'une expertise, d'un travail préparant une décision urgente ou à long terme ? Est-il utile d'intégrer les aspects biologiques, chimiques, économiques,... ou peut-on se limiter au aspects hydrodynamiques ? Une première simplification de la réalité est donc apportée par *sectorisation, i.e.* la limitation à un sous-modèle écologique, physique, économique, ...

Le nombre de variables d'état peut également être réduit par *agrégation, i.e.* en se limitant aux caractéristiques globales d'ensembles de variables d'état qui forment des *compartiments*. L'introduction du concept de salinité constitue un bel exemple d'agrégation puisque la salinité décrit en réalité la concentration d'un nombre important de sels dissous. De même, on peut s'intéresser à l'azote global dans le système, sans tenir compte des différentes formes sous lesquelles cet azote est présent (nitrite, nitrate, ammonium, N_2, matière organique,...). Dans les modèles d'écosystèmes, on intègre souvent les différentes espèces de producteurs primaires en un ou plusieurs groupes phytoplanctoniques. Dans ce cas, l'aggrégation peut également être réalisée d'un point de vue fonctionnel et non pas biologique ou physiologique.

D'un point de vue mathématique, l'agrégation de plusieurs variables d'état revient généralement à définir des moyennes pondérées de ces variables et de leurs équations d'évolution. Ainsi, définissons la concentration de l'agrégat par

$$C_{ag} = \sum_{i=1}^{N} \alpha_i C^i \tag{7.75}$$

où C^i sont les concentrations des espèces de base et α_i des constantes définissant le poids respectif de ces différentes espèces dans l'agrégat. La dynamique de chacune des espèces de base est décrite par une équation du type

$$\frac{\partial C^i}{\partial t} + \nabla \cdot \left(C^i (\mathbf{v} + \mathbf{v}^i_{s/m}) \right) - \nabla \cdot \left(\lambda \nabla C^i \right) - Q^i_c = 0 \tag{7.76}$$

Calculant la moyenne pondérée de ces équations comme dans (7.75), il vient

$$\sum_{i=1}^{N} \alpha_i \left[\frac{\partial C^i}{\partial t} + \nabla \cdot \left(C^i \mathbf{v} \right) - \nabla \cdot \left(\lambda \nabla C^i \right) \right] = \sum_{i=1}^{N} \alpha_i \left[Q^i_c - \nabla \cdot (C^i \mathbf{v}^i_{s/m}) \right] \tag{7.77}$$

Les différents termes du membre de gauche ne posent aucun problème. Ils s'expriment aisément en terme de la nouvelle variable C_{ag} :

$$\sum_{i=1}^{N} \alpha_i \left[\frac{\partial C^i}{\partial t} + \nabla \cdot \left(C^i \mathbf{v} \right) - \nabla \cdot \left(\lambda \nabla C^i \right) \right] = \frac{\partial C_{ag}}{\partial t} + \nabla \cdot (C_{ag} \mathbf{v}) - \nabla \cdot \left(\lambda \nabla C_{ag} \right) \tag{7.78}$$

Par contre, les termes placés dans le membre de droite font toujours apparaître les concentrations partielles C^i. Afin de réduire la portée du modèle, il est indispensable de

developer des paramétrisations appropriées de ces termes qui ne font apparaître que les seules variables relatives aux compartiments retenus dans le modèle afin de pouvoir écrire

$$\frac{\partial C_{ag}}{\partial t} + \nabla \cdot (C_{ag}\mathbf{v}) - \nabla \cdot \left(\lambda \nabla C_{ag}\right) = Q_{ag} - \nabla \cdot (C_{ag}\mathbf{v}^{ag}_{s/m}) \tag{7.79}$$

Alors que les termes de production-destruction Q^i_c et de migration/sédimentation $\mathbf{v}^i_{s/m}$ des différentes espèces prises individuellement peuvent généralement être paramétrisés en se basant sur des résultats expérimentaux simples, il n'en va plus de même pour les termes correspondants de l'agrégat. Les taux d'interactions correspondant ne sont en effet plus directement accessibles à l'expérience ou ne sont pas des caractéristiques biologiques ou chimiques intrinsèques. Le terme de migration/sédimentation constitue un exemple simple de cette difficulté. Alors, que les particules en suspension de même nature peuvent généralement être caractérisées par des vitesses de sédimentation $\mathbf{v}^i_{s/m}$ bien définies dépendant de leur taille, l'agrégat constitué des particules en suspension sans distinction de tailles peut difficilement être caractérisé par une vitesse unique $\mathbf{v}^{ag}_{s/m}$. Il en va de même de la modélisation des interactions.

Remarquons que les niveaux intermédiaires d'agrégation sont les plus sensibles à ces problèmes de paramétrisation. Lorsque le niveau d'agrégation est très élevé, la dynamique globale est généralement plus simple et, avec elle, les problèmes de paramétrisation.

7.5.4 Fenêtre spectrale.

De même qu'il est nécessaire de réduire la portée d'un modèle par sectorisation et agrégation, il est également nécessaire de choisir une échelle de temps.

Que ce soit en raison de leur dynamique propre ou des forçages auxquels ils sont soumis, les systèmes réels sont généralement le siège des processus caractérisés par des échelles de temps très différentes qu'il est impossible de décrire par un seul modèle. Aucun modèle ne peut représenter le spectre de fréquence tout entier, des processus moléculaires aux bouleversements climatiques, de la biologie des micro-organismes à l'écologie de l'environnement global.

La multiplicité des échelles temporelles est fortement influencée par la dynamique non-linéaire des systèmes. Les systèmes linéaires sont en général caractérisé par un petit nombre de temps caractéristiques intrinsèques. De plus, leur réponse à une excitation extérieure possède exactement la même fréquence que la sollicitation. Dans le cas d'un système non linéaire, il en va tout autrement. En effet, les non-linéarités entraînent des échanges entre des processus d'échelles de temps (et d'espace) différentes accompagnés de transfert énergétiques animant toutes les composantes du spectre. Parallèlement, les forçages internes et externes agissant sur le système tendent, comme dans le cas linéaire, à intensifier des composantes particulières dans des domaines d'échelles correspondant aux leurs. Dès lors, le spectre d'une variable d'état d'un système non linéaire est naturellement constitué d'une succession de pics et de vallées.

La figure 7.6 présente une vue schématique des différentes échelles de temps caractérisant les processus physiques en milieu marin. Comme on peut le voir, les

processus couvrent une vaste gamme de temps caractéristiques, de moins d'une seconde à l'échelle climatique, de la diffusion moléculaire à la circulation océanique profonde.

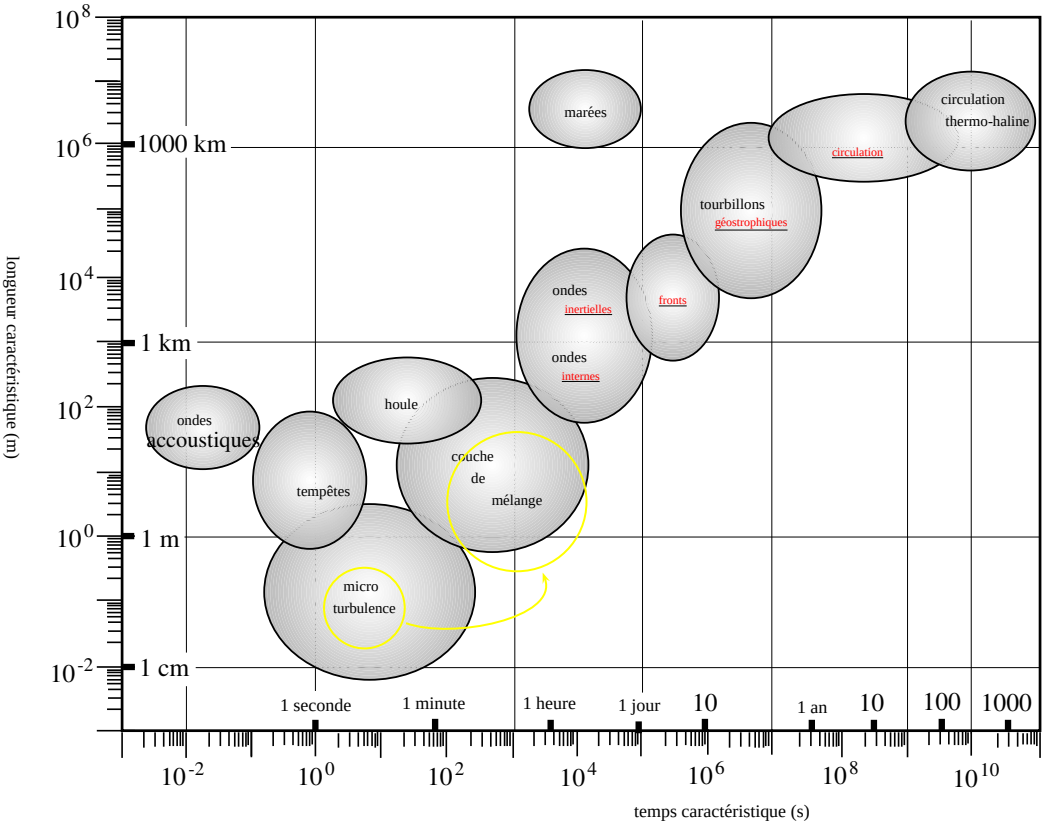


FIG. 7.6 – Échelles caractéristiques de temps et d'espace de la variabilité marine.

Comme le montre l'équation générale (7.74), l'hydrodynamique influence la dynamique des substances dissoutes et en suspension directement par le biais des processus d'advection et de diffusion (et indirectement par le contrôle de la distribution de la température et de la salinité). Les échelles de temps des processus hydrodynamiques se retrouvent donc également au niveau des processus biologiques et chimiques. Plus exactement, les processus biologiques et chimiques interagissent avec les processus hydrodynamiques qui possèdent des temps caractéristiques proches l'un de l'autre.

Un modèle particulier ne peut espérer représenter qu'une gamme particulière d'échelles temporelles, *i.e.* une *fenêtre spectrale*, qui définit l'*ouverture* du modèle. Bien évidemment, le choix d'une telle fenêtre doit toujours correspondre aux événements intenses associés au pic du spectre correspondant à l'objectif du modèle. Dans ce cadre, les variables d'état sont également spécifiées par leur échelle caractéristique de temps et constituent des moyennes représentatives d'une fenêtre spectrale. Elles peuvent

être définies mathématiquement comme des moyennes temporelles glissantes effectuées sur une période de temps T correspondant à la vallée spectrale en aval immédiat du pic étudié. Grace à cette opération de moyenne, les phénomènes à plus petite échelle que la fenêtre spectrale retenue ne sont pas résolus par le modèle mais éliminés par filtrage. Les phénomènes dont les temps caractéristiques excèdent la limite supérieure de la fenêtre spectrale retenue, ne seront pas non plus modélisés mais sont sous-jacents dans la définition du contexte général du problème et des conditions initiales. Enfin, seuls les processus appartenant à la fenêtre spectrale retenue sont modélisés et décrits explicitement.

L'équation générale (7.74) est a priori valable pour décrire toutes les échelles de temps. Pour examiner l'effet de la définition des variables opérantes du modèle par filtrage, décomposons les différentes variables selon

$$y = y' + \bar{y} \tag{7.80}$$

où $\bar{y} = < y >$ désigne la variable opérante dans la fenêtre spectrale retenue et y' représente la fluctuation à plus haute fréquence superposée. On a donc

$$< y' > = 0, \qquad < \bar{y} > = \bar{y} \tag{7.81}$$

Appliquant l'opérateur de moyenne $< \ldots >$ à chacun des termes de (7.74), on obtient

$$\frac{\partial \bar{C}}{\partial t} + \nabla \cdot \left(\bar{C}\bar{\mathbf{v}} + < C'\mathbf{v}' > \right) = \nabla \cdot \left(\lambda \nabla \bar{C} \right) + < Q_c > - \nabla \cdot \left(< C\mathbf{v}_{s/m} > \right) \tag{7.82}$$

i.e. les fluctuations disparaissent de tous les termes linéaires mais pas des termes non linéaires.

Comme dans le cas de l'agrégation, la paramétrisation des termes de production-destruction (et, dans une moindre mesure, celle de la sédimentation/migration) doit être adaptée pour faire apparaître uniquement les variables moyennes $\bar{y}$. Par exemple, dans le cas d'un processus qui dépend fortement du cycle jour-nuit, comme beaucoup de processus biologiques, il peut être très difficile de filtrer ces oscillations et de dégager une paramétrisation appropriée s'appuyant sur des variables moyennes définies par un temps caractéristique bien supérieur à 24 heures.

Sans entrer dans le détail de la paramétrisation de ces termes, le terme d'advection illustre la complexité de cette procédure. En effet, la moyenne du terme non linéaire d'advection donne naissance à deux termes dans (7.82) :

$$< C\mathbf{v} > = \bar{C}\bar{\mathbf{v}} + < C'\mathbf{v}' > \tag{7.83}$$

Le premier terme correspond au produit des valeurs moyennes et se trouve sous une forme appropriée. Le second terme, par contre, fait apparaître exclusivement les fluctuations. Cette moyenne du produit des fluctuations peut jouer un rôle important dès qu'il existe une corrélation significative entre les variables. Dans ce cas, il ne peut être ignoré mais il ne peut pas non plus être intégré tel quel au modèle opérant dans la fenêtre spectrale choisie.

Pour obtenir une paramétrisation appropriée de cette moyenne du produit de fluctuations, on émet généralement l'hypothèse que ces termes contribuent à l'uniformisation spatiale des propriétés, comme le processus de diffusion moléculaire dont λ tient compte. Dès lors, on introduit une paramétrisation semblable à celle de la diffusion moléculaire, *i.e.*

$$< C'\mathbf{v}' >= -\tilde{\lambda}\nabla\bar{C} \tag{7.84}$$

où $\tilde{\lambda}$ désigne le *coefficient de diffusion turbulente*. Le mécanisme de cette diffusion est essentiellement gouvernée par l'écoulement lui-même : le mélange est effectué par les tourbillons aux échelles temporelles inférieures à la fenêtre spectrale choisie. En raison de la taille importante de ces tourbillons, le mélange est beaucoup plus efficace et rapide que le mélange par diffusion moléculaire. L'intensité du mélange, et avec elle la valeur de $\tilde{\lambda}$, dépend essentiellement de l'énergie associée à ces tourbillons non résolus et est identique pour toutes les substances. On écrit généralement

$$\tilde{\lambda} \propto \ell\sqrt{k} \tag{7.85}$$

où k désigne l'énergie cinétique des fluctuations turbulentes et ℓ la longueur caractéristique des plus grands tourbillons non résolus.

Il faut remarquer que, alors que la diffusion moléculaire est isotrope, *i.e.* présente la même efficacité dans toutes les directions de l'espace, la diffusion turbulente devient anisotrope à partir d'une certaine échelle. D'une part, en effet, les mers et océans présentent des longueurs caractéristiques horizontales bien supérieures aux longueurs caractéristiques verticales (limitées par la profondeur). D'autre part, la stratification inhibe les échanges verticaux. Il en résulte que les tourbillons océaniques sont généralement plutôt bidimensionnels que tridimensionnels. Dès lors, la diffusion selon les directions horizontales et verticale est caractérisée par des intensités différentes et des coefficients de diffusions différents. On écrira donc

$$< C'\mathbf{v}' >= -\tilde{\lambda}_h\nabla_h\bar{C} - \tilde{\lambda}_v\frac{\partial C}{\partial z} \tag{7.86}$$

où $\tilde{\lambda}_h$ et $\tilde{\lambda}_v$ désignent les coefficients de diffusion horizontale et verticale et

$$\nabla_h = \frac{\partial}{\partial x}\mathbf{e}_x + \frac{\partial}{\partial y}\mathbf{e}_y \tag{7.87}$$

désigne la partie horizontale de l'opérateur ∇. De même, la diffusion verticale étant inhibée par la stratification, on écrira

$$\tilde{\lambda}_v = \psi(R_i)\ell\sqrt{k} \tag{7.88}$$

où $\psi(R_i)$ est une fonction décroissante du nombre de Richardson mesurant la stratification, *i.e.*

$$R_i = \frac{\overline{\frac{\partial b}{\partial z}}}{\overline{\frac{\partial \mathbf{v}}{\partial z}} \cdot \overline{\frac{\partial \mathbf{v}}{\partial z}}} = \frac{N^2}{M^2} \tag{7.89}$$

où b désigne la poussée, N la fréquence de Brunt-Väisälä et M la fréquence de Prandtl.

Adoptant la paramétrisation décrite ci-dessous et interprétant toutes les variables comme des moyennes correspondant aux variables opérantes du modèle dans la fenêtre spectrale choisie (en laissant tomber les barres), la forme générale de l'équation gouvernant la dynamique des substances dissoutes et en suspension est donnée par

$$\frac{\partial C}{\partial t} + \nabla \cdot \left[(\mathbf{v} + \mathbf{v}_{s/m}) C \right] = \nabla_h \cdot \left(\tilde{\lambda}_h \nabla_h C \right) + \frac{\partial}{\partial z} \left(\tilde{\lambda}_v \frac{\partial C}{\partial z} \right) + Q_c \tag{7.90}$$

Ce modèle séparant les dimensions horizontale et verticale considère cependant une stratification purement horizontale. Il est physiquement plus réaliste de représenter le processus de diffusion par deux termes correspondant respectivement à la diffusion isopycnale, *i.e.* le long des surfaces d'égale densité, et diapycnale, *i.e.* au travers du gradient de densité. Une forme plus générale de (7.90) est donc donnée par

$$\frac{\partial C}{\partial t} + \nabla \cdot \left[(\mathbf{v} + \mathbf{v}_{s/m}) C \right] = \nabla \cdot \left(\mathbf{K} \cdot \nabla C \right) + Q_c \tag{7.91}$$

où $\mathbf{K}$ désigne un tenseur de diffusion.

Remarquons que cette paramétrisation générale de la diffusion est généralement appropriée pour les études décrivant explicitement les mésoéchelles (temps caractéristiques de quelques heures à quelques jours). Pour des modélisation à des échelles de temps plus grandes, la résultante non linéaire des processus à mésoéchelle peut ne pas correspondre à une homogénéisation, même anisotrope, des propriétés mais induire une structure spatiale bien définie.

7.5.5 Intégration spatiale.

À partir du modèle tridimensionnel général (7.91) de l'équation d'évolution, des modèles simplifiés peuvent être construits en intégrant cette équation selon une ou plusieurs dimensions spatiales.

Modèle 0D.

L'intégration de (7.91) sur un volume V selon les trois dimensions de l'espace conduit à un *modèle boîte* ou *modèle 0D* ne permettant pas de décrire les variations spatiales des grandeurs étudiées. En effet, les variables d'un tel modèle sont définies, à partir des grandeurs variables spatialement, par

$$M_C = \iiint_V C dV \tag{7.92}$$

et représentent donc la masse totale comprise dans le volume V considéré.

Intégrant chacun des termes de (7.91) sur le volume V, il vient

$$\frac{dM_C}{dt} = \iiint_V Q_c dV + \iiint_V \left(\nabla \cdot \left(\mathbf{K} \cdot \nabla C - \nabla \cdot \left[(\mathbf{v} + \mathbf{v}_{s/m}) C \right] \right) \right) dV \tag{7.93}$$

Transformant la dernière intégrale en intégrales de flux par le théorème de Gauss, on obtient

$$\frac{dM_C}{dt} = \iiint_V Q_c dV + \iiint_\Sigma \left(\mathbf{K} \cdot \nabla C - (\mathbf{v} + \mathbf{v}_{s/m})C\right) \cdot \mathbf{n} \; d\Sigma \tag{7.94}$$

Cette expression constitue l'équation de bilan pour le constituent C : les variations temporelles de la masse de C sont dues aux interactions dans le domaine considéré et aux échanges au travers de la frontière Σ (frontières latérales, surface et fond).

Les intégrales de flux apparaissant dans le membre de droite de (7.94) correspondent aux conditions aux limites du modèle 3D initial. Ainsi donc, par intégration spatiale, les conditions aux limites ne forment plus des conditions auxiliaires du problème principal mais sont totalement intégrées à l'équation différentielle ordinaire décrivant la dynamique de M_C.

Remarquons que les termes de flux de (7.94) dépendent de la concentration C à la frontière du domaine étudié alors que la variable d'état M_C ne permet pas d'en décrire les variations spatiales. Il est donc nécessaire de développer une paramétrisation des échanges (sauf pour les flux qui ne dépendent pas explicitement de la concentration C au sein du domaine étudié) au travers de la frontière Σ en fonction de M_C. Cette paramétrisation constitue nécessairement une approximation des flux réels. En général, faute de mieux, on suppose que la distribution de C est uniforme au sein du volume V et que la valeur à la frontière Σ est donc donnée par M_C/V.

La même hypothèse d'uniformité (ou celle d'une distribution spatiale fixée et indépendante du temps) est nécessaire pour paramétriser les termes d'interaction. Cette hypothèse est évidemment très restrictive et ne s'applique pas à un grand nombre de problèmes d'interaction où le taux de production/destruction dépend cruellement de la probabilité de rencontre de deux espèces (chimiques ou biologiques).

Modèle 1D.

Trois types de modèles unidimensionnels sont possibles.

- Lorsqu'on étudie une rivière, on peut être tenté d'intégrer les équation sur la section de la rivière. Les variables d'état du modèle sont alors les concentrations moyennes sur ces sections. Elles varient en fonction de la coordonnée longitudinale selon la direction d'écoulement de la rivière.
- L'intégration selon les deux coordonnées horizontales, conservant donc la dimension verticale, est particulièrement bien adaptée à la modélisation des écosystèmes marins. De nombreux processus (dont la photosynthèse) dépendent en effet de l'éclairement, lequel varie principalement avec la profondeur, et/ou de la température. L'évolution de la thermocline constitue généralement un élément capital de la dynamique.
- Dans certaines études globales de l'océan, les grandeurs caractéristiques sont intégrées sur la profondeur et sur la longitude pour mettre en évidence les variations en fonction de la latitude.

L'expression particulière de ces modèles peut être obtenue en intégrant l'expression générale (7.91). Dans tous les cas, cette intégration introduit des termes de flux supplémentaires dans les équations et demande une paramétrisation des flux et interactions en fonction de grandeurs moyennes.

7.5.6 Modèle 2D.

Deux versions différentes de modèles 2D peuvent être envisagées.

- Dans la première, les équations 3D initiales initiales sont intégrées selon une dimension horizontale. Les variables d'état dépendent alors explicitement du temps, d'une coordonnée spatiale horizontale et de la profondeur. Elles s'interprètent alors comme des moyennes sur une 'tranche verticale' d'océan. De tels modèles sont utiles dans des études de processus où une direction horizontale peut être privilégiée. C'est par exemple le cas, dans des modèles simplifiés du talus continental ou d'autres zones d'upwelling.
- L'intégration des équations 3D selon la coordonnée verticale a été très souvent servi de base à la construction des premiers modèles hydrodynamiques des mers continentales. La dynamique des marées et des tempêtes peut en effet être décrite avec une très haute fidélité au moyen de tels modèles. Ici, les variables doivent être considérées comme des valeurs moyennes sur la colonne d'eau.

  Une approche semblable est suivie dans certains modèles biologiques où les équations 3D sont intégrées sur la couche de mélange en s'appuyant sur l'homogénéisation rapide des propriétés dans cette couche.

Considérons ici les aspects mathématiques de l'intégration verticale sur toute la colonne d'eau.

La frontière supérieure du domaine est matérialisée par la surface libre de la mer. Mesurant l'élévation de cette surface libre par rapport à un niveau de référence, l'équation de cette surface s'écrit sous la forme

$$\zeta(x,y,t) - z = 0 \tag{7.95}$$

De même, l'équation du fond s'écrit

$$h(x,y) + z = 0 \tag{7.96}$$

Dans ces deux relations, on suppose que la coordonnée verticale z est mesurée positivement vers le haut. Remarquons que le fond est ici considéré indépendant du temps (au contraire de la surface libre qui évolue au cours du temps). C'est évidemment une bonne approximation pour beaucoup de processus et pour des échelles de temps inférieures à l'échelle des mouvements sédimentaires ou géologiques.

La surface et le fond constitue des frontières naturelles du fluide que celui-ci ne peut traverser. On en déduit que la dérivée matérielle des relations (7.95) et (7.96) est nulle,

soit

$$\frac{\partial \zeta}{\partial t} + u_s \frac{\partial \zeta}{\partial x} + v_s \frac{\partial \zeta}{\partial y} - w_s = 0 \tag{7.97}$$

$$u_f \frac{\partial h}{\partial x} + v_f \frac{\partial h}{\partial y} + w_f = 0 \tag{7.98}$$

où (u_s, v_s, w_s) et (u_f, v_f, w_f) désignent les trois composantes de la vitesse respectivement à la surface et au fond. Géométriquement, la condition (7.98) impose que la vitesse du fluide est parallèle au fond. La condition (7.97) peut être interprétée de façon semblable à la surface mais intègre, en plus, l'effet des variations temporelles de la hauteur d'eau.

Considérons tout d'abord l'équation de continuité (7.66), *i.e.*

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \tag{7.99}$$

(cas particulier de l'équation générale (7.90) avec $C = 1$) et intégrons celle-ci selon la coordonnée verticale. Il vient

$$\int_{-h}^{\zeta} \frac{\partial u}{\partial x} dz + \int_{-h}^{\zeta} \frac{\partial v}{\partial y} dz + w_s - w_f = 0 \tag{7.100}$$

Les deux intégrales peuvent être exprimées en fonction des transports intégrés sur la profondeur

$$U(x,y,t) = \int_{-h(x,y)}^{\zeta(x,y,t)} u(x,y,z,t) dz, \qquad V(x,y,t) = \int_{-h(x,y)}^{\zeta(x,y,t)} v(x,y,z,t) dz \tag{7.101}$$

D'après la formule (1.75), on a

$$\begin{aligned} \frac{\partial}{\partial x} U(x,y,t) = \int_{-h(x,y)}^{\zeta(x,y,t)} \frac{\partial u}{\partial x}(x,y,z,t) dz \\ + \frac{\partial \zeta}{\partial x} u(x,y,\zeta(x,y,t)) + \frac{\partial h}{\partial x} u(x,y,-h(x,y,t)) \end{aligned} \tag{7.102}$$

Dès lors

$$\int_{-h}^{\zeta} \frac{\partial u}{\partial x}(x,y,z,t) dz = \frac{\partial U}{\partial x} - u_s \frac{\partial \zeta}{\partial x} - u_f \frac{\partial h}{\partial x} \tag{7.103}$$

$$\int_{-h}^{\zeta} \frac{\partial v}{\partial x}(x,y,z,t) dz = \frac{\partial V}{\partial y} - v_s \frac{\partial \zeta}{\partial y} - v_f \frac{\partial h}{\partial y} \tag{7.104}$$

Substituant ces relations dans (7.100), on obtient

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \left[w_s - u_s \frac{\partial \zeta}{\partial x} - v_s \frac{\partial \zeta}{\partial y} \right] - \left[w_f + u_f \frac{\partial h}{\partial x} + v_f \frac{\partial h}{\partial y} \right] = 0 \tag{7.105}$$

Tenant compte des conditions aux limites (7.97) et (7.98) et introduisant la notation

$$\mathbf{U} = U\mathbf{e}_x + V\mathbf{e}_y, \qquad H(x,y,t) = \zeta(x,y,t) + h(x,y) \tag{7.106}$$

Il vient finalement

$$\frac{\partial H}{\partial t} + \nabla_h \cdot \mathbf{U} = 0 \tag{7.107}$$

qui constitue la forme bidimensionnelle intégrée sur la profondeur de l'équation de continuité ; puisque l'eau de mer est considérée incompressible, toute divergence du transport horizontal doit être compensée par une augmentation de la hauteur de la colonne d'eau.

Appliquons le même traitement à l'équation (7.90). Nous définissons la concentration moyenne sur la colonne d'eau par

$$\bar{C}(x,y,t) = \frac{1}{\zeta + h}\int_{-h}^{\zeta} C(x,y,z,t)dz \tag{7.108}$$

Intégrons ensuite les différents termes de (7.90) (en négligeant les termes de migration-sédimentation).

– La dérivée temporelle de la concentration donne naissance à deux termes :

$$\int_{-h}^{\zeta} \frac{\partial C}{\partial t}dz = \frac{\partial}{\partial t}\int_{-h}^{\zeta} Cdz - C_s\frac{\partial \zeta}{\partial t} = \frac{\partial (H\bar{C})}{\partial t} - C_s\frac{\partial \zeta}{\partial t} \tag{7.109}$$

où C_s désigne la concentration en surface.

– La partie horizontale du terme d'advection se transforme selon

$$\int_{-h}^{\zeta} \nabla_h \cdot (\mathbf{u}C)dz = \nabla_h \cdot \int_{-h}^{\zeta} \mathbf{u}Cdz - \mathbf{u}_s C_s \cdot \nabla_h \zeta - \mathbf{u}_f C_f \cdot \nabla_h h \tag{7.110}$$

où C_f désigne la concentration en $z = -h$.

– La partie verticale de l'advection peut s'écrire sous la forme

$$\int_{-h}^{\zeta} \frac{\partial}{\partial z}(wC)dz = \Big[wC\Big]_{z=-h}^{z=\zeta} = w_s C_s - w_f C_f \tag{7.111}$$

– Le terme décrivant la diffusion horizontale s'écrit

$$\int_{-h}^{\zeta} \nabla_h \cdot \left(\tilde{\lambda}_h \nabla_h C\right) dz = \nabla_h \cdot \int_{-h}^{\zeta} \tilde{\lambda}_h \nabla_h C dz - \tilde{\lambda}_h \nabla_h C\Big|_{z=\zeta} \cdot \nabla_h \zeta - \tilde{\lambda}_h \nabla_h C\Big|_{z=-h} \cdot \nabla_h h \tag{7.112}$$

– Le terme de diffusion verticale devient

$$\int_{-h}^{\zeta} \frac{\partial}{\partial z}\left(\tilde{\lambda}_v \frac{\partial C}{\partial z}\right) dz = \left[\tilde{\lambda}_v \frac{\partial C}{\partial z}\right]_{z=-h}^{z=\zeta} = J_f^{diff} - J_s^{diff} \tag{7.113}$$

où J_f^{diff} et J_s^{diff} désignent les flux de diffusion en surface et au fond.

– Le terme de production-destruction peut s'écrire sous la forme

$$\int_{-h}^{\zeta} Q_c dz = H\bar{Q_c} \tag{7.114}$$

où $\bar{Q}$ désigne le taux moyen de production-destruction sur la colonne d'eau.

En regroupant les termes de (7.109)-(7.111), les termes de surface et de fond se simplifient grâce à (7.97) et (7.98). On a

$$\int_{-h}^{\zeta} \left(\frac{\partial C}{\partial t} + \nabla \cdot (\mathbf{v}C) \right) dz = \frac{\partial (H\bar{C})}{\partial t} + \nabla_h \cdot \int_{-h}^{\zeta} \mathbf{u}C dz \tag{7.115}$$

Pour adapter cette forme au modèle 2D, le second terme du membre de droite doit être exprimé en fonction de $\bar{C}$, H et $\mathbf{U}$. Pour ce faire, décomposons chacune des variables 3D en sa moyenne sur la profondeur et sa déviation par rapport à cette moyenne (notée par $'$), *i.e.*

$$C = \bar{C} + C', \qquad \mathbf{u} = \frac{\mathbf{U}}{H} + \mathbf{u}' \tag{7.116}$$

Il vient alors

$$\int_{-h}^{\zeta} \mathbf{u}C dz = \mathbf{U}\bar{C} + \int_{-h}^{\zeta} \mathbf{u}'C' dz \tag{7.117}$$

Comme dans le cas de l'intégration temporelle, le second terme de cette expression ne peut être exprimé en fonction des variables moyennes mais doit être paramétrisé en fonction de celle-ci. On suppose généralement que ce terme augmente la diffusion horizontale des propriétés moyennes du fluide ; on parle alors de diffusion par cisaillement ('shear effect diffusion'). On introduit donc la paramétrisation

$$\int_{-h}^{\zeta} \mathbf{u}'C' dz = -K^{shear} H \nabla \bar{C} \tag{7.118}$$

Les différents termes relatifs à la diffusion horizontale (7.112) ne font pas non plus apparaître explicitement les variables moyennes. Une nouvelle modélisation du processus de diffusion en terme de ces variables s'impose donc selon[5]

$$-\int_{-h}^{\zeta} \tilde{\lambda}_h \nabla_h C dz = -K^{diff} H \nabla \bar{C} \tag{7.119}$$

qui se combine avec (7.118) pour former un terme de diffusion unique combinant les deux processus.

Si on parvient à exprimer les flux, à la surface et au fond, ainsi que le terme de production $\bar{Q_c}$ en terme de la concentration moyenne $\bar{C}$, le modèle bidimensionnel complet s'écrit donc

$$\frac{\partial (H\bar{C})}{\partial t} + \nabla_h \cdot \left(\mathbf{U}\bar{C} \right) = \nabla_h \cdot \left(HK^{tot} \nabla_h \bar{C} \right) + H\bar{Q}_C + J_f^{diff} - J_s^{diff} \tag{7.120}$$

[5]Remarquons qu'il n'est pas correct de modéliser la diffusion par un terme proportionnel à $\nabla \left(H\bar{C} \right)$ puisque celui-ci tendrait à distribuer la concentration moyenne en raison inverse de la profondeur locale.

où $K^{tot} = K^{shear} + K^{diff}$.

Utilisant la forme bidimensionnelle de l'équation de continuité (7.107), on peut également écrire (en adoptant une paramétrisation légèrement différente de la diffusion),

$$\frac{\partial \bar{C}}{\partial t} + \frac{\mathbf{U}}{H} \cdot \nabla_h \bar{C} = \nabla_h \cdot \left(K^{tot} \nabla_h \bar{C} \right) + \bar{Q}_C + \frac{J_f^{diff} - J_s^{diff}}{H} \tag{7.121}$$

Cette expression montre que l'influence des flux en surface et au fond est inversement proportionnelle à la profondeur.

7.5.7 Filtrage spatial.

Comme le suggère la figure 7.6, les échelles de temps et d'espace des processus hydrodynamiques sont liées entre-elles. Les processus les plus rapides sont aussi ceux qui possèdent les plus petites échelles de temps. Inversement, les processus les plus lents sont caractérisés par de très grandes longueurs. Dès lors, tout filtrage temporel tel qu'à la section 7.5.4 induit simultanément un filtrage spatial ; les processus aux plus petites échelles d'espace sont automatiquement éliminés des équations. Ainsi, si on se concentre sur la dynamique à mésoéchelle (temps caractéristiques de quelques heures à quelques jours), on devra interpréter les variables opérantes du modèle non seulement comme des moyennes sur un temps caractéristiques d'une dizaine de minutes mais aussi comme des moyennes sur des longueurs caractéristiques horizontales de quelques dizaines de mètres.

De ce qui précède, on pourrait conclure qu'il n'est pas nécessaire de considérer explicitement le filtrage spatial dès lors qu'un filtrage temporel a été effectué. Un filtrage spatial est cependant introduit indépendamment du filtrage temporel lors de la résolution numérique des modèles résolus spatialement. En effet, la discrétisation spatiale introduite pour résoudre numériquement les équations du type de (7.91) induit un filtrage spatial indépendent du choix de la fenêtre spectrale. Les variables d'un modèle numérique (nécessairement de résolution finie) doivent être considérées comme des moyennes sur les mailles de la grille numérique. Cette re-définition des variables opérantes doit s'accompagner, comme dans le cas du filtrage temporel, d'une modélisation des effets non linéaires des processus sous-grille qui ne peuvent être décrits explicitement par le modèle numérique. Faute de mieux, la modélisation choisie est, une fois encore, celle d'un terme supplémentaire de diffusion. (Celui-ci est généralement actif uniquement selon l'horizontale puisque la résolution verticale est suffisante pour la fenêtre spectrale retenue.) En anticipant la résolution numérique des équations, ce terme est souvent inclus dans le modèle mathématique bien qu'il trouve sa justification dans la considération des aspects numériques. Le coefficient de diffusion correspondant dépend a priori de la fenêtre spectrale du modèle et du pas spatiale de la grille numérique.

7.5.8 Ajustement écohydrodynamique.

La présentation ci-dessus à fait la part belle aux processus hydrodynamiques. Il y a une bonne raison pour cela.

Les processus biologiques et chimiques possèdent également des temps caractéristiques propres.

Ainsi, par exemple, les populations phytoplanctoniques possèdent des temps caractéristiques de 10^5 à 10^7 secondes correspondant à des cycles de vie caractéristiques de beaucoup de populations pélagiques et benthiques (oscillations diurnes, saisonnières, annuelles). À de telles échelles de temps, le phytoplancton est fortement influencé par la marée et la circulation générale saisonnière, *i.e.* par les processus hydrodynamiques (regroupés sous l'appellation de 'temps de la mer') possèdant des temps caractéristiques semblables.

Clairement, les processus hydrodynamiques dont les temps caractéristiques sont plus petits (resp. plus grands) que les temps caractéristiques biologiques ne peuvent interagir efficacement avec ceux-ci. En général, seuls les processus hydrodynamiques dont les temps caractéristiques sont proches de ceux des interactions biogéochimiques peuvent avoir une influence et contraindre la dynamique biologique et chimique. Par le biais de l'advection et de la diffusion, ces processus hydrodynamiques impriment alors leurs longueurs caractéristiques aux variables biologiques et chimiques. C'est ce que l'on appelle l'*ajustement écohydrodynamique.*

Chapitre 8

Modélisation au moyen de nuages de particules.

8.1 Modélisation Lagrangienne.

En écrivant l'équation de bilan (7.90), nous avons représenté l'état du système par la concentration de quelques variables caractéristiques. Le concept de concentration ainsi utilisé est lié à celui de milieu continu. Selon cette approche, pour décrire le système marin, on suppose que l'on peut définir des grandeurs locales variant d'un point à l'autre du système. C'est ainsi, par exemple, que l'on caractérise la distribution de la masse par unité de volume ρ. Celle-ci est calculée en considérant le rapport entre la masse m_Δ d'un échantillon et le volume Δ de cet échantillon. Pour obtenir une grandeur locale, on souhaiterait calculer

$$\lim_{\Delta \to 0} \frac{m_\Delta}{\Delta} \tag{8.1}$$

À proprement parler, cette limite n'a pas de sens. En effet, dès lors que l'on descend à l'échelle moléculaire, voire atomique, le rapport m_Δ/Δ présente un comportement erratique et ne possède pas de limite au sens mathématique du terme.

Dans l'approche des milieux continus, on définit donc la masse par unité de volume ρ par

$$\rho = \frac{m_\Delta}{\Delta} \tag{8.2}$$

où Δ désigne un ensemble de molécules suffisamment grand pour donner un sens statistique aux moyennes et aux grandeurs résultantes comme m_Δ et suffisamment petit pour capter les variations spatiales de ces grandeurs aux échelles qui nous intéressent. On décrit donc le milieu par un ensemble de propriétés variant (quasi-)continûment en faisant fi des discontinuités aux niveaux moléculaire et atomique.

L'approche du milieu continu est appropriée pour décrire les champs de vitesse, de température, de concentration de nutriments…Elle peut aussi être utilisée pour représenter la distribution du plancton ou même celle des populations de poissons, à

condition que l'on s'intéresse à des longueurs caractéristiques bien supérieures à celle des individus formant ces populations.

Le concept de milieu continu ne s'applique par contre pas à la représentation et à l'étude de la dynamique d'individus pris isolément. Si, par exemple, on peut étudier la dynamique des populations de harengs en mer du Nord en décrivant celle-ci au moyen de sa concentration, il n'est pas possible de procéder de la même façon pour étudier le comportement d'un individu dans un banc. Pour cette dernière étude, il est nécessaire de pouvoir distinguer l'individu dans le groupe. De même, pour étudier et modéliser la migration des baleines, on préférera suivre chaque individu spécifiquement au cours de son périple. Cette approche donne lieu à des modèles 'individus centrés' (ou 'Individual Base Model - IBM' en anglais).

Une approche semblable est possible lors de l'étude de la dispersion de polluants. Ainsi, la plupart des modèles de dispersion d'hydrocarbures, une nappe de polluant est représentée comme un ensembles de particules dont on suit les évolutions séparées au cours du temps.

Les modèles individus centrés et les modèles de dispersion d'hydrocarbures évoqués ci-dessus correspondent à une description Lagrangienne du système : les variables du modèle sont attachées à chaque individu ou chaque particule dont on suit l'évolution au cours du temps. Cette approche s'oppose à l'approche Eulérienne développée au chapitre précédent et qui consiste en la description des variations spatiales et temporelles des grandeurs caractéristiques. Là où l'approche Eulérienne décrit les variations temporelles en un point fixe, l'approche Lagrangienne décrit l'évolution temporelle pour une particule donnée en suivant celle-ci au cours de son mouvement. Ces deux approches se retrouvent au niveau des méthodes expérimentales d'observation : les lignes de mouillages décrivent les courants et les propriétés de l'eau en un endroit fixe alors que les bouées dérivantes en fournissent une description Lagrangienne.

8.2 Modélisation Lagrangienne de l'advection et de la diffusion.

La trajectoire de chaque particule ou de chaque individu peut être décrite par la donnée de sa position $\mathbf{s}(t)$ à chaque instant à partir d'un instant initial t_0, soit, composante par composante

$$(X(t), Y(t), Z(t)), \qquad t \geq t_0 \tag{8.3}$$

La représentation de l'advection d'une particule par le fluide en mouvement ne pose pas de problème. Si la particule se déplace à la vitesse $\mathbf{v}$, il vient simplement

$$\begin{aligned} \mathbf{s}(t+\Delta t) &= \mathbf{s}(t) + \int_t^{t+\Delta t} \mathbf{v}\ dt' \\ &\sim \mathbf{s}(t) + \mathbf{v}\Delta t + o(\Delta t) \end{aligned} \tag{8.4}$$

Dans la majorité des cas, la vitesse $\mathbf{v}$ des particules est la vitesse du fluide. Comme celle-ci est généralement donnée dans un formalisme Eulérien selon $\mathbf{v} = \mathbf{v}(\mathbf{s},t)$, l'équation (8.4) s'écrit dans ce cas

$$\begin{aligned} \mathbf{s}(t+\Delta t) &= \mathbf{s}(t) + \int_t^{t+\Delta t} \mathbf{v}(\mathbf{s}(t'),t') \ dt' \\ &\sim \mathbf{s}(t) + \mathbf{v}(\mathbf{s}(t),t)\Delta t + o(\Delta t) \end{aligned} \tag{8.5}$$

Les processus de sédimentation et la migration peuvent également être aisément pris en compte en introduisant l'expression appropriée de la vitesse dans (8.4).

La description de la diffusion est autrement plus complexe. Celle-ci introduit en effet un caractère non déterministe dans la dynamique individuelle des particules de sorte que les résultats doivent être interprétés de façon statistique.

Macroscopiquement ou à l'échelle d'une population d'individus, la diffusion se traduit par le lissage des variations spatiales des variables d'état. À l'échelle des tourbillons à l'origine de la diffusion turbulente ou à l'échelle des individus, la turbulence se marque par des mouvements erratiques semblables au mouvement Brownien. Dès lors, deux particules occupant une même position à l'instant initial suivront des trajectoires différentes au cours du temps. Pour obtenir une description statistiquement significative des caractéristiques globales du système, il est donc nécessaire de réaliser des simulations avec un très grand nombre de particules.

Le principe de la représentation Lagrangienne de la diffusion est d'ajouter un déplacement aléatoire $\mathbf{d}$ (variable) au déplacement déterministe associé à l'advection selon

$$\mathbf{s}(t+\Delta t) = \mathbf{s}(t) + \int_t^{t+\Delta t} \mathbf{v} \ dt' + \mathbf{d} \tag{8.6}$$

Notant $\mathbf{s}^i = \mathbf{s}(t_0 + i\Delta t)$ et utilisant une valeur approchée de l'intégrale, on aura

$$\mathbf{s}^{i+1} = \mathbf{s}^i + \mathbf{v}\Delta t + \mathbf{d}^i \tag{8.7}$$

Dans le cas de la diffusion moléculaire isotrope, les sauts successifs $\mathbf{d}^i$ sont statistiquement indépendants et reliés au coefficient de diffusion habituelle K par le biais de

$$\mathbf{d}^i = \boldsymbol{\delta}\sqrt{2K\Delta t} \tag{8.8}$$

où $\boldsymbol{\delta}$ désigne un vecteur aléatoire issu d'une distribution Gaussienne normalisée.

L'expression (8.8) est pleinement justifiée dans le cas où le coefficient de diffusion est constant et uniforme. Dans ce cas, elle conduit, avec (8.7), à une augmentation de la variance de la position varie proportionnelle au nombre de pas de temps (ou au temps total écoulé) et au coefficient de diffusion K comme le modèle Eulérien correspondant.

D'une part, on vérifie aisément que la solution du problème unidimensionnel de diffusion

$$\frac{\partial C}{\partial t} = K\frac{\partial^2 C}{\partial t^2} \tag{8.9}$$

dans un milieu infini est en effet donnée par

$$C(t,z) = \frac{1}{\sqrt{4\pi Kt}} \exp\Big[-\frac{x^2}{4Kt}\Big] \tag{8.10}$$

La variance de cette distribution est donnée par

$$\sigma^2 = 2Kt \tag{8.11}$$

D'autre part, considérons le mouvement d'une particule décrit par le processus stochastique

$$x_0, \qquad x_{k+1} = x_k + \tilde{d}_k L, \qquad k = 0,1,2,\ldots \tag{8.12}$$

où $\tilde{d}_k = \pm 1$ est choisi de façon aléatoire (avec une probabilité égale pour les deux issues +1 et -1). On calcule aisément que, après N itérations,

$$< (x_N)^2 >= NL^2 \tag{8.13}$$

soit, si les N itérations correspondent à N incréments temporels Δt,

$$< (x_N)^2 >= \frac{tL^2}{\Delta t} \tag{8.14}$$

En comparant les deux expression (8.11) et (8.14), on en déduit que les deux (8.9) et (8.12) décrivent le même phénomène de diffusion pour autant que

$$L = \sqrt{2K\Delta t} \tag{8.15}$$

De plus, par le théorème de la valeur centrale, la distribution binomiale utilisée dans (8.12) tend vers une distribution normale ce qui justifie pleinement (8.8).

Dans le cas de la diffusion turbulente, variant dans le temps et l'espace, la formule (8.8) n'est plus strictement équivalente à la modélisation Eulérienne correspondante. Elle est cependant généralement employée pour obtenir séparément les différentes composantes de $\mathbf{d}^i$ en remplaçant le coefficient de diffusion par le coefficient de diffusion turbulente dans la direction correspondante.

8.3 Modélisation individu-centrée.

Pour construire un modèle Lagrangien décrivant les aspects biologiques ou chimiques d'un système, il suffit d'attacher à chaque particule des caractéristiques propres, des règles d'évolution au cours du temps et des règles d'interaction avec les autres particules. L'advection et la diffusion sont alors décrites comme dans la section précédente alors que la dynamique propre et les interactions sont gouvernées par ces nouvelles règles.

L'un des aspects particulièrement attrayants de la modélisation individu-centrée est la possibilité de tenir compte explicitement de la variabilité individuelle en considérant que les caractéristiques des différents individus sont extraites de distributions statistiques

pré-définies. Alors que ses caractéristiques sont généralement réduites à une seule valeur moyenne pour toute la population dans un modèle Eulérien, il est donc possible de tenir compte des différences entre individus et de tenir compte de l'influence de ces différences sur la dynamique du système.

La construction d'un modèle individu-centré passe par trois étapes importantes :

i. Avant tout, il convient de dresser la liste des individus ou groupes d'individus dont on désire suivre l'évolution et d'associer, à chaque individu, les caractéristiques pertinentes pour l'étude envisagée : position dans l'espace, âge, taille, état nutritionnel, ...

ii. Les individus sont destinés à évoluer dans un certain environnement dont il faut définir les caractéristiques : topographie/bathymétrie, conditions hydrodynamiques et physiques, forçages,...

iii. Enfin il faut définir les règles qui régissent la dynamique de chaque individu. Outre les règles de mise à jour de la position pour tenir compte de l'advection et de la diffusion, on décrire ainsi, par exemple, les règles qui conditionnent les migrations verticales et horizontales, la stratégie de sélection des proies, la reproduction, ...

En pratique, la mise en place d'un modèle individu-centré ne demande aucune technique mathématique ou numérique complexe. la simplicité de l'approche de base est pour beaucoup dans l'intérêt porté à ces modèles. Cependant, l'analyse de la dynamique d'un tel modèle est généralement heuristique. L'approche la plus courante consiste à multiplier les expériences numériques en variant les paramètres et paramétrisations dans le but d'identifier des conclusions globales. Dans les meilleurs des cas, cependant les conclusions sont essentiellement statistiques. Elle permettent d'identifier des corrélations entre grandeurs globales caractéristiques de l'ensemble de la population mais fournissent rarement une explication mécanistique de la dynamique du système étudié.

8.4 Comparaison entre les modèles Eulérien et Lagrangien.

À cause du grand nombre de particules à prendre en compte dans les modèles Lagrangiens, ceux-ci peuvent être beaucoup plus coûteux que les modèles Eulériens correspondants pour la description de grandes zones géographiques.

Si on désire décrire la distribution d'un polluant sur le Plateau Continental Nord Ouest Européen, un modèle Eulérien demandera la résolution d'une seule équation semblable à (7.90) pour la concentration de ce polluant. La résolution d'une telle équation est généralement réalisée sur une grille numérique couvrant le domaine considéré avec une résolution spatiale d'une dizaine de kilomètres. Pour obtenir la même résolution spatiale sur l'ensemble du plateau continental avec un modèle Lagrangien, on devra avoir de l'ordre de quelques centaines de particules par maille de la grille Eulérienne ! Le coût informatique correspondant est absolument prohibitif.

En présence d'un rejet local, par contre, l'approche Lagrangienne est particulièrement bien adaptée. En effet, alors que le traitement Eulérien du problème requière la résolution d'une équation du type de (7.90) dans toute la zone couverte par le modèle, l'approche Lagrangienne permet de se concentrer uniquement sur la zone réellement affectée par le rejet où se trouvent un nombre important de particules.

Les biologistes habitués à décrire la physiologie des individus sont spontanément attirés par l'approche Lagrangienne. Celle-ci permet, par exemple, une description plus naturelle de la succession des stades des copépodes ou des comportements individuels. Un grand niveau de complexité, et donc de réalisme, peut être atteint dans la modélisation Lagrangienne des comportements individuels tout en suivant une approche simple et instinctive. Pour obtenir une description globale d'une population, cependant, il faut simuler le devenir d'un très grand nombre d'individus interagissant entre-eux pour obtenir des résultats statistiquement significatifs. L'approche Eulérienne, représentant globalement la dynamique d'une population peut alors se révéler préférable, même si elle demande une paramétrisation artificielle des effets des interactions (non-linéaires) entre individus sur l'ensemble de la population.

Ceci suggère une approche intermédiaire dans laquelle la dynamique des populations est décrite par un modèle Eulérien dans lequel les paramétrisations des interactions entre individus sont développées en se basant sur des modèles de processus de type individu-centré.

Annexe A

Exercices proposés

A.1 Équations différentielles

1) $y' = \dfrac{\sin^2 x}{\sin y}$, Rép. : $2\cos y - \sin x \cos x + x = C$

2) $2\sqrt{y} = y'$, Rép. : $y = (x+C)^2$

3) $1 + y^2 + xyy' = 0$, Rép. : $x^2(1+y^2) = C$

4) $(1+e^x)yy' = e^x$, Rép. : $y^2 = 2\,\ln(1+e^x) + C$

5) $y'' - 2y' = e^x \sin x$,

Rép. : $y = C_1 + C_2 e^{2x} - \dfrac{1}{2} e^x \sin x$

6) $y'' - 7y' + 6y = \sin x$,

Rép. : $y = C_1 e^x + C_2 e^{6x} + \dfrac{1}{74}(5\sin x + 7\cos x)$

7) $y' = \dfrac{4y^2}{x^2} - y^2$

Rép. : $y = \dfrac{x}{x^2 + Cx + 4}$

8) Un circuit électrique est constitué de la mise en série d'une résistance R, d'un condensateur C et d'une force électromotrice variable $V(t)$. La charge q du condensateur obéit dès lors à l'équation

$$R\,\frac{dq}{dt} + \frac{q}{C} = V(t)$$

Déterminez la charge $q(t)$ si $V(t) = V_0 \sin \omega t$ et si $q(0) = 0$.

Rép. : $q(t) = \dfrac{C V_0}{1 + \omega^2 R^2 C^2}\,(\sin \omega t + RC\omega[\exp(-t/RC) - \cos \omega t])$

9) Un flotteur plongé dans l'eau subit son propre poids et la poussée d'Archimède (égale au poids du liquide déplacé). Si la section A du flotteur est constante, la hauteur immergée $z(t)$ du flotteur varie selon

$$m\frac{d^2z}{dt^2} = mg - \rho Agz$$

où m est la masse du flotteur, g l'accélérateur de pesanteur et ρ la masse volumique de l'eau.

Déterminez le comportement du flotteur lorsqu'on le dépose sans vitesse juste à la surface du liquide.

Rép. : $z(t) = \dfrac{m}{\rho A}\left(1 - \cos\sqrt{\dfrac{\rho Ag}{m}}\,t\right)$

10) Lorsqu'un corps de température absolue T est plongé dans un environnement à la température différente T_{ext} (supposée constante), sa température s'adapte progressivement à celle de son environnement. Selon la loi de Newton, le flux de chaleur échangé entre ce corps et son environnement est proportionnel à la différence de température $T - T_{ext}$. La température du corps varie donc selon

$$\frac{d}{dt}T(t) = -\alpha(T(t) - T_{ext})$$

où α est une constante positive. Déterminez $T(t)$ si initialement $T(0) = T_0 \neq T_{ext}$.

Rép. : $T(t) = T_{ext} + (T_0 - T_{ext})\,\mathrm{e}^{-\alpha t}$

Le modèle de Newton s'applique lorsque l'échange de chaleur se produit essentiellement par conduction thermique. Si les échanges radiatifs dominent, les flux de chaleur émis par le corps et par son environnement sont proportionnels à la quatrième puissance de la température absolue (loi de Stefan-Boltzman). Dès lors, on aura

$$\frac{d}{dt}T(t) = -\beta(T^4(t) - T_{ext}^4)$$

où β est une constante positive. Comment varie la température du corps dans ce cas ?

Rép. : $4\beta T_{ext}^3 t = \ln\dfrac{(T + T_{ext})(T_0 - T_{ext})}{(T - T_{ext})(T_0 + T_{ext})} + 2\,\mathrm{arctg}\,\dfrac{T}{T_{ext}} - 2\,\mathrm{arctg}\,\dfrac{T_0}{T_{ext}}$

11) La vitesse à laquelle un fluide s'écoule d'un réservoir est proportionnelle à la racine carrée de la hauteur de liquide au-dessus de l'orifice.

– Dans le cas d'un réservoir cylindrique de section (horizontale) constante A percé d'un orifice à sa base, déterminez le niveau du liquide dans le réservoir en fonction du temps sachant que le réservoir se vide en un temps T.

Rép. : $y(t) = y_0(1 - t/T)^2$

– Idem dans le cas d'un réservoir conique de hauteur H et de rayon maximum R.

Rép. : $y(t) = y_0(1-t/T)^{2/5}$

12) La résistance exercée par l'air sur un corps en chute libre est souvent supposée proportionnelle au carré de la vitesse du corps. Sous cette hypothèse, déterminez la vitesse d'un corps en chute libre en résolvant

$$\frac{dv}{dt} = +g - kv^2 \quad (v(t) \geq 0)$$

et en considérant que la vitesse initiale du corps est nulle.

Rép. : $v(t) = \sqrt{\frac{g}{k}} \frac{e^{2t\sqrt{gk}} - 1}{e^{2t\sqrt{gk}} + 1}$

Déterminez la hauteur du corps en fonction du temps si $y(0) = y_0$

Rép. : $y(t) = y_0 - \sqrt{\frac{g}{k}} t + \frac{1}{k} \ln\left(\frac{1 + e^{2t\sqrt{gk}}}{2}\right)$

13) On considère un circuit électrique composé d'une force électromotrice V constante, d'une résistance R et d'une self L placées en série. A l'instant initial, le circuit n'est parcouru par aucun courant. On a

$$L\frac{di}{dt} + Ri = V$$

– Que vaut $\lim_{t\to\infty} i(t)$?

Rép. : $i_\infty = \frac{V}{R}$

– Après combien de temps le courant électrique atteint-il 90 % de sa valeur limite ?

Rép. : $\tau = \ln 10 \frac{L}{R}$

14) Résolvez les systèmes suivants.

a) $\begin{cases} y_1' &= y_1 + y_2, \\ y_2' &= 4y_1 - 2y_2, \end{cases}$

Rép. : $\begin{cases} y_1 = C_1 e^{-3x} + C_2 e^{2x} \\ y_2 = -4C_1 e^{-3x} + C_2 e^{2x} \end{cases}$

b) $\begin{cases} y_1' &= 2y_1 - 5y_2 - \sin 2x, & y_1(0) = 0, \\ y_2' &= y_1 - 2y_2 + x, & y_2(0) = 1 \end{cases}$

Rép. : $\begin{cases} y_1 = -\frac{4}{3}\sin x + \frac{2}{3}\sin 2x - \frac{2}{3}\cos x + \frac{2}{3}\cos 2x - 5x \\ y_2 = -\frac{2}{3}\sin x + \frac{1}{3}\sin 2x - 2x + 1 \end{cases}$

c) $\begin{cases} y_1' &= y_1, \\ y_2' &= -y_2 + \sqrt{2}y_3, \\ y_3' &= \sqrt{2}y_2. \end{cases}$

Rép. : $\begin{cases} y_1 = C_2 e^x \\ y_2 = C_3 e^x - \sqrt{2}C_1 e^{-2x} \\ y_3 = \sqrt{2}C_3 e^x + C_1 e^{-2x} \end{cases}$

15) Un médecin arrivant sur le lieu d'un crime constate que la température du mort est de 32° et que la température de l'air ambient est de 18°C. Deux heures plus tard, la température du mort est descendue à 26°. En supposant que le taux de refroidissement du corps est proportionnel à la différence de température entre l'air et le corps de la victime (loi de Newton) et que la température du corps au moment du décès était de 36°C, déterminez le temps écoulé depuis la mort de la victime jusqu'à l'arrivée du médecin.

Rép. : 54 minutes

16) Soit un circuit électrique RC-série dont la résistance varie au cours du temps selon $R = \alpha + \beta t$ où α et β sont des constantes positives. Si une différence de potentiel E constante est appliquée, la charge q du condensateur varie selon

$$Rq'(t) + \frac{1}{C}q(t) = E$$

Déterminez $q(t)$ si $q(0) = q_0$.

Rép. : $q(t) = EC + (q_0 - EC)\left(\dfrac{\alpha}{\alpha+\beta t}\right)^{1/C\beta}$

17) Soit une goutte d'eau tombant du ciel. En supposant que la goutte reste parfaitement sphérique,

a) déterminez le rayon $r(t)$ de la goutte à chaque instant sachant que le taux d'évaporation de l'eau est proportionnel à la puissance α de la surface de la goutte ;

Rép. : $\begin{cases} \left(R_0^{3-2\alpha} - 2^{-2+2\alpha}\pi^{-1+\alpha}(3-2\alpha)\beta t\right)^{\frac{1}{3-2\alpha}} & \text{si } \alpha \neq 3/2 \\ R_0\, e^{-2\sqrt{\pi}\beta t} & \text{si } \alpha = 3/2 \end{cases}$

b) déterminez les valeurs de α compatibles avec l'évaporation totale de la goutte en un temps fini ;

Rép. : $\alpha < 3/2$

18) Lorsqu'un élément subit une décroissance radioactive, celui-ci se transforme en un autre élément qui peut lui-même être radioactif. les concentrations x et y des ces

deux éléments sont alors données par

$$\begin{cases} \dot{x}(t) = -\lambda_1 x(t) \\ \dot{y}(t) = \lambda_1 x(t) - \lambda_2 y(t) \end{cases}$$

Déterminez les concentrations des éléments si $x(0) = x_0$ et $y(0) = 0$.

Rép. : $x(t) = x_0\,\mathrm{e}^{-\lambda_1 t}, y(t) = x_0\lambda_1 \dfrac{\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\lambda_2 t}}{\lambda_2 - \lambda_1}$

19) Si on tient compte des pertes de mémoires, le taux de mémorisation d'un cours est donné par

$$\frac{dA}{dt}(t) = \alpha(M - A(t)) - \beta A(t)$$

où α et β sont des constantes positives, où $A(t)$ désigne la quantité de matière mémorisée et où M désigne la quantité de matière totale à mémoriser.

– Déterminez la quantité de matière mémorisée lorsque $t \to \infty$.
– Déterminez $A(t)$ si $A(0) = 0$.

Rép. : $A_\infty = \alpha M/(\alpha+\beta),\ A(t) = A_\infty(1 - \mathrm{e}^{-(\alpha+\beta)t})$

20) Si un médicament est administré en continu via une perfusion, la concentration $x(t)$ de ce médicament dans le sang est gouvernée par l'équation

$$\frac{dx}{dt}(t) = \alpha - \beta x(t)$$

où α et β sont des constantes positives.

– Déterminez la concentration du médicament dans le sang lorsque $t \to \infty$.
– Déterminez $x(t)$ si $x(0) = 0$.

Rép. : $x_\infty = \alpha/\beta,\ x(t) = x_\infty(1 - \mathrm{e}^{-\beta t})$

21) Si la croissance d'une certaine espèce de poissons suit une loi logistique et si un nombre constant de poissons est pêché chaque année, la dynamique de la population de cette espèce peut être représentée par l'équation

$$\frac{dP}{dt}(t) = P(t)(\alpha - \beta P(t)) - \gamma$$

où α, β et γ sont des constantes positives.

Dans le cas où $\alpha = 5$, $\beta = 1$, $\gamma = 4$, $P(0) = P_0$,

– déterminez $P(t)$;

Rép. : $P(t) = \dfrac{4(P_0 - 1) - (P_0 - 4)\,\mathrm{e}^{-3t}}{(P_0 - 1) - (P_0 - 4)\,\mathrm{e}^{-3t}}$

– montrez que, dans certaines conditions, l'espèce sera complètement éteinte en un temps fini t° (Le modèle n'est pas applicable au-delà de cet instant.).

Rép. : Si $P_0 < 1,\ t^\circ = \dfrac{1}{3}\ln\dfrac{P_0 - 4}{4P_0 - 4}$

22) On considère deux réactifs A et B qui se combinent pour former une troisième espèce chimique C. La réaction est telle que, pour chaque gramme de A impliqué dans la réaction, 4 grammes de B sont utilisés. On observe que 30 grammes de C se sont formés en 10 minutes. Déterminez la quantité de C présente à chaque instant sachant que la vitesse de la réaction est proportionnelle aux produit des concentrations des deux réactifs. À l'instant initial, il n'y a pas de C mais seulement 50 grammes de A et 32 grammes de B.

Rép. : $C(t) = 1000\dfrac{1-\mathrm{e}^{-0.1258t}}{25-4\,\mathrm{e}^{-0.1258t}}$

23) On considère une vasque hémisphérique de rayon R=10 mètres. Initialement, la vasque est vide. Un débit de π m^3/s est amené pour remplir la vasque. Sachant que le taux d'évaporation est donné par $0.01A(t)$ où $A(t)$ désigne l'aire de la surface libre, déterminez l'évolution du niveau de l'eau dans la vasque. La vasque pourra-t-elle être remplie ?

Remarque : Pour une hauteur d'eau dans la vasque égale à h, le volume est donné par $\pi Rh^2 - \pi h^3/3$.

Rép. : $h_\infty = 10m$

24) L'équation de la déformation d'une poutre élastique supportant une charge uniformément répartie sur toute sa longueur ℓ est donnée par

$$EI\frac{d^4y}{dx^4} = w_0 \qquad 0 \le x \le \ell$$

où EI représente la rigidité flexionnelle de la poutre et où w_0 représente la charge par unité de longueur.

Déterminez la déformation d'une poutre encastrée à ses deux extrémités, c'est-à-dire telle que

$$y(0) = y(\ell) = 0, \qquad y'(0) = y'(\ell) = 0$$

Rép. : $y(x) = \dfrac{w_0}{24EI}x^2(x-\ell)^2$

25) La distribution de la température $T(r)$ dans la région comprise entre deux sphères concentriques de rayons $r = a$ et $r = b$ $(a < b)$ est gouvernée par

$$r\frac{d^2T}{dr^2} + 2\frac{dT}{dr} = 0 \qquad \text{où } T(a) = T_0, \quad T(b) = T_1$$

Déterminez $T(r)$.

Rép. : $\dfrac{T_0 - T_1}{b-a}\dfrac{ab}{r} + \dfrac{T_1 b - T_0 a}{b-a}$

26) La distribution de la température $T(r)$ dans la région comprise entre deux cylindres concentriques de rayons $r = a$ et $r = b$ $(a < b)$ est gouvernée par

$$r\frac{d^2T}{dr^2} + \frac{dT}{dr} = 0 \qquad \text{où } T(a) = T_0, \quad T(b) = T_1$$

Déterminez $T(r)$.

Rép. : $\dfrac{T_0 \ln r/b - T_1 \ln r/a}{\ln a/b}$

A.2 Équations aux différences

1) Trouvez la solution pour chacune des relations de récurrence suivantes données avec leur condition initiale.
   a) $a_n = 3a_{n-1}, a_0 = 2$ Rép. : 2.3^n
   b) $a_n = a_{n-1} + 2, a_0 = 3$ Rép. : $3 + 2n$
   c) $a_n = a_{n-1} + n, a_0 = 1$ Rép. : $1 + \dfrac{n(n+1)}{2}$
   d) $a_n = a_{n-1} + 2n + 3, a_0 = 4$ Rép. : $(2+n)^2$
   e) $a_n = 2a_{n-1} - 1, a_0 = 1$ Rép. : 1
   f) $a_n = 3a_{n-1} + 1, a_0 = 1$ Rép. : $\dfrac{1}{2}(3^{n+1} - 1)$
   g) $a_n = na_{n-1}, a_0 = 5$ Rép. : $5n!$
   h) $a_n = 2na_{n-1}, a_0 = 1$ Rép. : $2^n n!$
2) Résolvez les relations de récurrence suivantes avec les conditions initiales données.
   a) $a_n = 2a_{n-1}$ pour $n \geq 1, a_0 = 3$ Rép. : $3\,2^n$
   b) $a_n = a_{n-1}$ pour $n \geq 1, a_0 = 2$ Rép. : 2
   c) $a_n = 5a_{n-1} - 6a_{n-2}$ pour $n \geq 2, a_0 = 1, a_1 = 0$ Rép. : $3\,2^n - 2\,3^n$
   d) $a_n = 4a_{n-1} - 4a_{n-2}$ pour $n \geq 2, a_0 = 6, a_1 = 8$ Rép. : $2^{n+1}(3-n)$
   e) $a_n = -4a_{n-1} - 4a_{n-2}$ pour $n \geq 2, a_0 = 0, a_1 = 1$ Rép. : $(-2)^{n-1} n$
   f) $a_n = 4a_{n-2}$ pour $n \geq 2, a_0 = 0, a_1 = 4$
   Rép. : $2^n - (-2)^n$ soit $\begin{cases} 2^{n+1} \text{ si } n \text{ impair} \\ 0 \text{ si } n \text{ pair} \end{cases}$
   g) $a_n = a_{n-2}/4$ pour $n \geq 2, a_0 = 1, a_1 = 0$
   Rép. : $\dfrac{1}{2}[(\dfrac{1}{2})^n + (-\dfrac{1}{2})^n]$ soit $\begin{cases} \dfrac{1}{2^n} \text{ si } n \text{ pair} \\ 0 \text{ si } n \text{ impair} \end{cases}$
   h) $a_n = 2(a_{n-1} - a_{n-2})$ pour $n \geq 2, a_0 = 1, a_1 = 2$
   Rép. : $(\sqrt{2})^n \left[\cos\dfrac{n\pi}{4} + \sin\dfrac{n\pi}{4}\right]$
   i) $a_n = -a_{n-2}$ pour $n \geq 2, a_0 = 0, a_1 = 3$
   Rép. : $3\sin\dfrac{n\pi}{2}$
   j) $a_n = -2a_{n-1} - 2a_{n-2}$ pour $n \geq 2, a_0 = 1, a_1 = 3$
   Rép. : $(\sqrt{2})^n \left[\cos\dfrac{3n\pi}{4} + 4\sin\dfrac{3n\pi}{4}\right]$
3) Résolvez les relations de récurrence suivantes avec les conditions initiales données.

a) $a_n = a_{n-1} + 6a_{n-2}$ pour $n \geq 2, a_0 = 3, a_1 = 6$ Rép. : $\frac{3}{5}((-2)^n + 4.3^n)$

b) $a_n = 7a_{n-1} - 10a_{n-2}$ pour $n \geq 2, a_0 = 2, a_1 = 1$ Rép. : $3.2^n - 5^n$

c) $a_n = 6a_{n-1} - 8a_{n-2}$ pour $n \geq 2, a_0 = 4, a_1 = 10$ Rép. : $3.2^n + 4^n$

d) $a_n = 2a_{n-1} - a_{n-2}$ pour $n \geq 2, a_0 = 4, a_1 = 1$ Rép. : $4 - 3n$

e) $a_n = a_{n-2}$ pour $n \geq 2, a_0 = 5, a_1 = -1$ Rép. : $2 + 3(-1)^n$

f) $a_n = -6a_{n-1} - 9a_{n-2}$ pour $n \geq 2, a_0 = 3, a_1 = -3$ Rép. : $(-3)^n(3 - 2n)$

g) $a_n = -4a_{n-1} + 56a_{n-2}$ pour $n \geq 0, a_0 = 2, a_1 = 8$

Rép. : $a_n = (-2)^n \Big[\left(1 + \sqrt{\frac{3}{5}}\right)(1 - \sqrt{15})^n + \left(1 - \sqrt{\frac{3}{5}}\right)(1 + \sqrt{15})^n \Big]$

4) Trouvez la solution de $a_n = 2a_{n-1} + a_{n-2} - 2a_{n-3}$ pour $n = 3, 4, 5, \ldots$ avec $a_0 = 3, a_1 = 6$ et $a_2 = 0$.

Rép. : $a_n = 6 - 2(-1)^n - 2^n$

5) Déterminez les valeurs des constantes A et B de manière telle que $a_n = An + B$ soit une solution de la relation de récurrence $a_n = 2a_{n-1} + n + 5$.

Rép. : $a_n = -n - 7$

6) Une personne dépose 1000 dollars sur un compte en banque qui rapporte un intérêt annuel de 9 %.

   a) Etablissez une relation de récurrence pour calculer le montant accumulé sur le compte à la fin de n années.

   b) Trouvez une formule explicite pour calculer le montant sur le compte à la fin de n années.

   c) Quelle est la valeur du compte après cent ans ?

Rép. : a) $a_n = a_{n-1}(1 + 0.09), \quad a_0 = 1000$; b) $a_n = 1000(1 + 0.09)^n$; c) 5529040

7) Supposez que la population mondiale en 1995 est de 7 milliards et qu'elle croît à raison de 3% par an.

   a) Établissez une relation de récurrence pour calculer la population mondiale dans n années après 1995.

   b) Trouvez une formule explicite pour calculer la population mondiale au bout de n années après 1995.

   c) Quelle sera la population mondiale en 2010 ?

Rép. : a) $p_n = p_{n-1}(1 + 0.03), \quad p_0 = 7.10^9$; b) $p_n = 7.10^9(1 + 0.03)^n$; c) $10.9058\,10^9$

8) Le mouvement erratique des particules au sein d'un fluide est appelé *mouvement brownien*. Celui-ci est responsable de la diffusion des propriétés du fluide dans tout son volume. On peut donner un modèle unidimensionnel de ce processus en considérant une particule qui, à chaque instant, possède une probabilité $p = 1/2$ de se déplacer d'une unité vers la droite et une probabilité $q = 1/2$ de se déplacer d'une

unité vers la gauche. Si on considère que les particules peuvent être absorbées par des parois situées en $x = 0$ et $x = \ell \in \mathbb{N}$, alors, la probabilité P_k pour qu'une particule située à l'abscisse $x = k \in \mathbb{N}$ soit absorbée par la paroi située en $x = 0$ est la solution de l'équation aux différences

$$P_k = pP_{k+1} + qP_{k-1} \quad (0 < k < \ell) \qquad \text{avec} \qquad P_0 = 1 \quad \text{et} \quad P_\ell = 0$$

Déterminez la probabilité P_k $(0 \leq k \leq \ell)$.

Rép. : $P_k = 1 - \dfrac{k}{l}$

9) Un modèle pour calculer le nombre de homards capturés par année est fondé sur l'hypothèse que le nombre de homards capturés dans une année est la moyenne du nombre capturé au cours des deux années précédentes.

   a) Trouvez une relation de récurrence pour $\{L_n\}$ où L_n est le nombre de homards capturés au cours de l'année n en tenant compte de l'hypothèse de ce modèle.
   Rép. : $L_n = \dfrac{1}{2}(L_{n-1} + L_{n-2}) \quad n \geq 3$

   b) Trouver L_n si 100 000 homards ont été capturés au cours de l'année 1 et 300 000 au cours de l'année 2.

Rép. : $L_1 = 100\,000, L_2 = 300\,000, L_n = \dfrac{100000}{3}\left(7 + \dfrac{(-1)^n}{2^{n-3}}\right)$

10) Tentant de résoudre numériquement le problème différentiel

$$\begin{aligned} \ddot{x} + 2\omega\dot{x} + \omega^2 x &= 0 \\ x(0) = 1,\ \dot{x}(0) &= 0 \end{aligned} \tag{A.1}$$

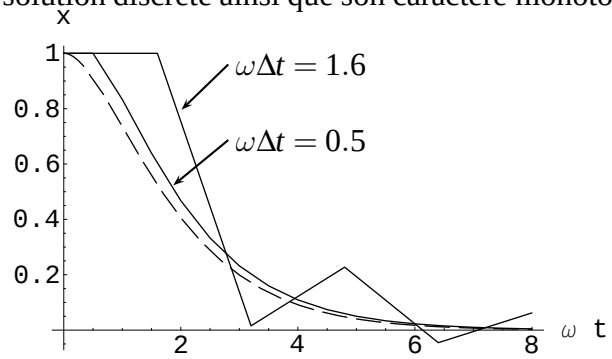
décrivant le comportement d'un oscillateur avec amortissement critique, on établit la discrétisation suivante

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{\Delta t^2} + \omega\frac{x_{k+1} - x_{k-1}}{\Delta t} + \omega^2 x_k = 0,\ x_0 = 1,\ x_1 = 1 \tag{A.2}$$

où x_k est la solution approchée au temps $t = k\Delta t$.

Les solutions de (A.2) pour différentes valeurs de Δt, soit $\omega\Delta t = 0.5$, $\omega\Delta t = 1.6$ et $\omega\Delta t = 2.1$ sont représentées ci-dessous ainsi que la solution du problème continu initial (A.1) (trait interrompu).

Expliquez le comportement de la solution discrète pour les différentes valeurs du pas d'intégration Δt données. En particulier, justifiez le caractère croissant ou décroissant de la solution discrète ainsi que son caractère monotone ou oscillatoire.



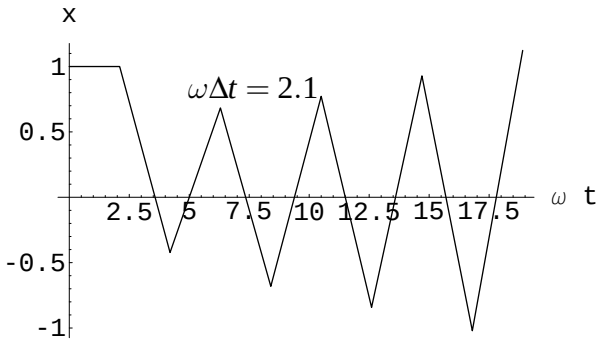

x
1
0.5
-0.5
-1
$\omega\Delta t = 2.1$
2.5
5
7.5
10
12.5
15
17.5
ω t

Table des matières