

AZÉDINE GRINE

QUELQUES MÉTHODES STATISTIQUES D'ESTIMATION DE
LA PRODUCTION D'ORGANISMES HÉTÉROTROPHES

Essai
présenté
à la Faculté des études supérieures
pour l'obtention
du grade de maître ès sciences (M. Sc.)

Département de mathématiques et de statistique
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL

AOÛT 2002

© Azédine Grine, 2002

RÉSUMÉ

En écologie, le calcul de la production des organismes hétérotrophes revêt une importance capitale. Dans la pratique, la détermination exacte de cette production dite secondaire n'est pas toujours possible. Cette difficulté a conduit au développement de plusieurs méthodes d'estimation, dont la majorité sont fondées sur un échantillonnage probabiliste de la population à différents moments dans le temps. On distingue trois grandes catégories de telles méthodes : a) les méthodes directes, au nombre de quatre, à savoir la sommation des accroissements, la sommation des pertes, la courbe de Allen et la technique du taux de croissance instantanée ; b) la méthode dite des fréquences de tailles ; c) les méthodes indirectes ou de régression. Cet essai présente les propriétés des méthodes directes d'estimation de la production secondaire, ainsi que celle des fréquences de tailles. On y présente aussi une étude de cas permettant de juger de la qualité de ces estimateurs dans le cadre d'une simulation portant sur une population fictive mais réaliste de moules bleues (*Mytilus edulis*).

Azédine Grine, étudiant

Christian Genest, directeur

REMERCIEMENTS

Mes plus vifs remerciements vont bien entendu à ma tendre épouse et à mes adorables enfants. La première pour m'avoir soutenu tout le temps et encouragé en permanence pour venir à bout de ce travail, les seconds pour avoir été une source inépuisable d'énergie et de motivation. Qu'ils trouvent ici le témoignage de mon profond amour.

Au plan scientifique, je tiens à remercier infiniment le Professeur Christian Genest pour avoir bien voulu me proposer ce sujet d'essai, ainsi que pour son encadrement et ses conseils très déterminants dans l'accomplissement de ce travail. Je suis également vivement reconnaissant envers M. Mathieu Cusson, du *Groupe interuniversitaire de recherches océanographiques du Québec*, pour toute l'aide qu'il m'a apportée dans la compréhension du "langage des biologistes," toute la bibliographie qu'il a mise à ma disposition, ainsi que pour sa très grande gentillesse et sa disponibilité. Je remercie également M. Jean-François Plante pour l'élaboration du programme informatique qui a rendu possible la simulation décrite dans le dernier chapitre du présent ouvrage.

Ce travail de recherche a été financé en partie par des octrois accordés au Professeur Genest par le *Conseil de recherches en sciences naturelles et en génie du Canada*, ainsi qu'à l'équipe de statistique de l'Université Laval par le *Fond québécois de recherche sur la nature et les technologies*.

Table des matières

RÉSUMÉ	ii
REMERCIEMENTS	iii
TABLE DES MATIÈRES	iv
LISTE DES TABLEAUX	vi
LISTE DES FIGURES	vii
INTRODUCTION	1
CHAPITRE I. NOTION ET CALCUL DE LA PRODUCTION	4
1.1 Définition de la production	4
1.2 Calcul de la production en général	5
1.3 La courbe de Allen	6
CHAPITRE II. MÉTHODES D'ESTIMATION DE LA PRO- DUCTION	9

2.1	Méthode de la sommation des accroissements	10
2.2	Méthode de la sommation des pertes	11
2.3	Méthode de la courbe de Allen	16
2.4	Méthode du taux de croissance instantanée	17
2.5	Méthode des fréquences de tailles	22
2.6	Résumé des méthodes	26
 CHAPITRE III. QUALITÉ DES ESTIMATEURS : UNE ÉTU-		
DE DE CAS		27
3.1	Travaux antérieurs	27
3.2	Description d'une population fictive	28
3.3	Description de l'expérience	33
3.4	Résultats	36
3.5	Discussion	45
 CONCLUSION		47
 BIBLIOGRAPHIE		49

LISTE DES TABLEAUX

Tableau 1a : Données de Crisp	14
Tableau 1b : Données de Crisp (suite)	15
Tableau 2 : Données de Waters et Crawford	22
Tableau 3 : Données de Rigler et Downing	25
Tableau 4 : Biais des estimateurs	43
Tableau 5 : Variance des estimateurs	44

LISTE DES FIGURES

Figure 1 : La relation entre N_t et m_t	7
Figure 2 : Surface entre m_i et m_{i+1}	10
Figure 3 : Surface entre N_i et N_{i+1}	12
Figure 4 : Logigramme	26
Figure 5 : Évolution de N_t en fonction du temps	31
Figure 6 : Évolution de m_t en fonction du temps	32
Figure 7 : Relation entre N_t et m_t	35
Figure 8 : Relation entre $\sqrt{N_t}$ et $\log(m_t)$	35
Figure 9 : Résultats pour la méthode de la sommation des accroissements	38
Figure 10 : Résultats pour la méthode de la courbe de Allen (version naïve)	39
Figure 11 : Résultats pour la méthode de la courbe de Allen (version sophistiquée)	40
Figure 12 : Résultats pour la méthode du taux de croissance instantanée (sans approximation)	41
Figure 13 : Résultats pour la méthode du taux de croissance instantanée (avec approximation)	42

INTRODUCTION

D'un point de vue fonctionnel, les écosystèmes peuvent être conceptualisés comme un ensemble de compartiments qui échangent de la matière et de l'énergie. Un des objectifs de l'écologie est de caractériser ces compartiments, dont l'importance peut être exprimée en terme de biomasse, ainsi que les flux qui se produisent entre eux.

Les flux en question, dont l'intensité est souvent exprimée en biomasse par unité de temps, peuvent correspondre, soit à des échanges entre deux compartiments différents (prédation, défécation, dégradation, etc.), soit à la production de nouvelle biomasse par un compartiment biologique donné. Cette production sera alors dite primaire ou secondaire suivant que le compartiment dont elle provient est autotrophe ou hétérotrophe.

La notion de production est fondamentale en écologie, puisque selon Benke (1973), c'est elle qui synthétise le mieux l'ensemble des données (biomasse, croissance individuelle, reproduction, temps de génération, survie, etc.) caractérisant le fonctionnement dynamique d'un compartiment biologique, d'une population ou d'un écosystème.

Cet essai porte sur les méthodes mathématiques et statistiques de mesure de la production. Celles-ci diffèrent très largement suivant la nature du compartiment biologique considéré. Les productions primaires et microbiennes sont uniquement appréhendées à l'échelle de la communauté dans son ensemble, et leur mesure fait largement utilisation de radiotraceurs. Les mesures de production secondaire des animaux macroscopiques sont, au contraire, le plus

souvent envisagées à l'échelle de la population. Elles sont réalisées directement à partir des mesures de densité et de masse individuelle et nécessitent un effort d'échantillonnage et d'analyse particulièrement important. Aussi ces dernières seront-elles ici l'objet exclusif de notre propos.

La détermination complète de la production secondaire est généralement impossible puisqu'elle dépend en grande partie :

- a) des caractéristiques (et plus particulièrement de la structure) de la population étudiée ;
- b) de l'effort et de la fréquence d'échantillonnage réalisés.

Cette difficulté a conduit au développement de plusieurs méthodes de mesure de la production secondaire. Classiquement, on distingue trois grandes catégories de telles méthodes :

- 1) les méthodes directes, au nombre de quatre, à savoir la sommation des accroissements, la sommation des pertes, la courbe de Allen et la technique du taux de croissance instantanée ;
- 2) la méthode des fréquences de tailles ;
- 3) les méthodes indirectes ou de régression.

Après avoir défini la notion de production et détaillé son calcul au chapitre 1 dans un cas de parfaite information, ce travail se concentre plus spécifiquement sur les méthodes directes d'estimation de la production, qui sont exposées et illustrées au chapitre 2. Une courte section à la fin de ce chapitre concerne en outre la technique dite des fréquences de tailles.

Étonnamment, on sait très peu de choses concernant les propriétés statistiques des méthodes directes de mesure de la production. À peine quelques études de simulation ont-elles été réalisées afin de déterminer le biais et la variance de ces estimateurs pour des populations biologiques possédant des caractéristiques précises. Le chapitre 3 cite quelques sources à cet effet et poursuit dans le même sens en présentant les résultats d'une petite étude de Monte-Carlo portant sur une population fictive de 1 000 000 d'individus ayant des taux de croissance et de mortalité s'apparentant à ceux de la moule bleue (*Mytilus edulis*). Comme on le verra, la fréquence et l'effort d'échantillonnage peuvent avoir un effet marqué sur la précision et la fiabilité des estimations obtenues.

CHAPITRE I

NOTION ET CALCUL DE LA PRODUCTION

1.1 Définition de la production

Considérons une population fermée composée de N individus dont les masses à l'instant t_D sont notées m_{1D}, \dots, m_{ND} . Supposons qu'en un temps $t_F > t_D$, ces individus soient encore vivants et que leurs masses soient alors égales à m_{1F}, \dots, m_{NF} . Si les individus sont en croissance, on s'attend à ce que $m_{iF} > m_{iD}$ pour tout $1 \leq i \leq N$, mais il peut arriver que $m_{iF} \leq m_{iD}$ pour certains sujets.

On appelle *production* de la cohorte sur l'intervalle $[t_D, t_F]$ et on note $P_{[t_D, t_F]}$ la somme sur tous les individus des écarts de masse $m_{iF} - m_{iD}$, $i = 1, \dots, N$. Autrement dit,

$$P_{[t_D, t_F]} = \sum_{i=1}^N (m_{iF} - m_{iD}).$$

Cette quantité est généralement positive, mais peut s'avérer négative ou nulle à l'occasion.

Observons que le calcul de $P_{[t_D, t_F]}$ ne nécessite pas la connaissance de toutes les valeurs m_{iD} et m_{iF} , mais seulement de leurs moyennes

$$m_D = \frac{1}{N} \sum_{i=1}^N m_{iD} \quad \text{et} \quad m_F = \frac{1}{N} \sum_{i=1}^N m_{iF}.$$

En effet, on a

$$P_{[t_D, t_F]} = N(m_F - m_D). \quad (1.1)$$

Pour connaître la production totale d'une population de taille N connue, il suffit donc d'en peser les membres *tous ensemble*, une première fois au temps t_D et une deuxième au temps t_F . Cette simplification est fort commode, puisqu'en pratique, il est généralement difficile d'identifier les membres d'une grande population et d'en déterminer la masse individuelle à deux moments précis dans le temps.

Les valeurs de Nm_D et de Nm_F , qui s'expriment généralement en kilogrammes, sont appelées *biomasses initiale et finale de la cohorte*.

1.2 Calcul de la production en général

Supposons maintenant que l'espèce concernée soit sujette à mortalité ou à prédation au cours de la période d'observation $[t_D, t_F]$, mais qu'aucune naissance n'intervienne dans ce même laps de temps. Il reste alors possible de déterminer la production de cette *cohorte* si on dispose, à chaque instant $t \in [t_D, t_F]$, des valeurs de

$$\begin{cases} N_t & = \text{le nombre d'individus vivants à l'instant } t \text{ et} \\ m_t & = \text{la moyenne de leurs masses individuelles,} \end{cases}$$

et donc de la biomasse $B_t = N_t m_t$ en tout point.

En effet, soient $t_1 < \dots < t_n$ les points de saut de la fonction N_t . Introduisons la notation

$$m_{t_i}^- = \lim_{t \uparrow t_i} m_t$$

pour représenter la masse individuelle moyenne, calculée sur l'ensemble des sujets vivants au temps t ou qui disparaissent (par mortalité ou prédation) à cet instant précis. En appliquant la formule (1.1) à chacun des sous-intervalles $[t_i, t_{i+1})$, on voit alors que la production totale de la cohorte sur l'intervalle $[t_D, t_F]$ est donnée par

$$P_{[t_D, t_F]} = \sum_{i=0}^n N_{t_i} \left(m_{t_{i+1}}^- - m_{t_i} \right), \quad (1.2)$$

où $t_0 = t_D$ et $t_{n+1} = t_F$.

Dans la pratique, la taille des populations étudiées est telle que les fonctions N_t et m_t peuvent être modélisées comme continues, voire dérivables en t . En particulier, les biologistes supposent généralement que $m_{t_i}^- = m_{t_i}$ pour tout t . Il est également commun de supposer que cette fonction est croissante sur l'ensemble de son domaine. C'est ce qui sera fait dans la suite.

1.3 La courbe de Allen

Allen (1950) a fait remarquer que la formule (1.2) correspond à l'aire d'un diagramme à bâtons dont les $n + 1$ colonnes auraient pour base $m_{t_{i+1}}^- - m_{t_i}$ et pour hauteur N_{t_i} , $i = 0, \dots, n$. En supposant que ces différences soient positives et que l'on ait $m_{t_i}^- = m_{t_i}$ pour tout i , les paires (m_{t_i}, N_{t_i}) forment alors une fonction décroissante en escalier (voir figure 1) dont l'intégrale représente la production totale sur l'intervalle $[t_D, t_F]$.

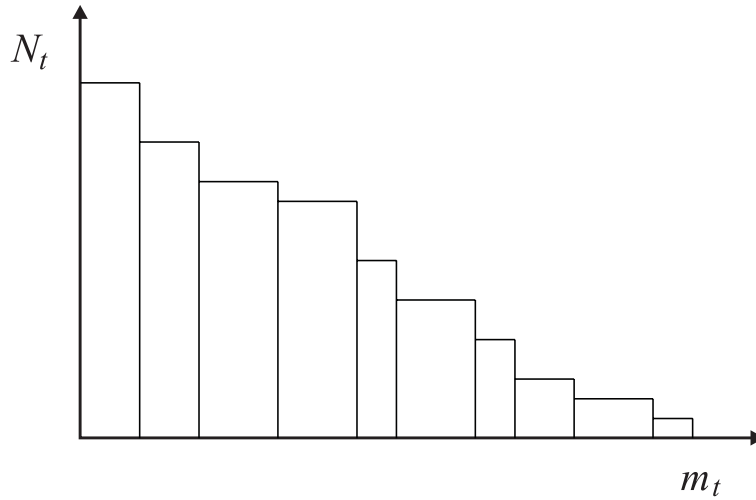


Figure 1 : La relation entre N_t et m_t . Cette illustration montre que le calcul d'une production est équivalent à la détermination de la surface d'un histogramme dont les fréquences sont les effectifs N_{t_i} et dont les longueurs de classe sont les différences $m_{t_{i+1}}^- - m_{t_i}$, $i = 0, \dots, n$. Le graphique présenté ici suppose que ces différences sont positives et que $m_{t_i}^- = m_{t_i}$ pour tout i .

De façon générale, on appelle *courbe de Allen* l'ensemble des points (m_t, N_t) obtenus en faisant varier t dans un intervalle $[t_D, t_F]$ donné. Tel qu'indiqué plus haut, il faut supposer la continuité et la croissance de m_t pour que cet objet mathématique ait un sens. Si on suppose en outre, par abstraction, la dérivabilité de N_t et de m_t , la biomasse $B_t = N_t m_t$ et la production P_t deviennent alors des fonctions lisses du temps. En particulier, le taux de production $P'_t = dP_t/dt$ représente alors l'accroissement instantané de la biomasse $B_t = N_t m_t$, une fois corrigée pour la perte due aux prédateurs ou autres causes de mortalité. De façon plus spécifique, on a

$$P'_t = B'_t - m_t N'_t = N_t m'_t + N'_t m_t - m_t N'_t = N_t m'_t, \quad (1.3)$$

d'où l'on déduit que

$$\begin{aligned} P_{[t_D, t_F]} &= \int_{t_D}^{t_F} N_t m'_t dt \\ &= \int_{t_D}^{t_F} B'_t dt - \int_{t_D}^{t_F} N'_t m_t dt \\ &= (B_{t_F} - B_{t_D}) - \int_{t_D}^{t_F} m_t dN_t. \end{aligned} \tag{1.4}$$

Notons que la formule (1.4) se réduit à (1.2) dans le cas spécial où N_t est constant par morceaux. En particulier, on retrouve la formule (1.1) si $N_t \equiv N$ partout sur l'intervalle $[t_D, t_F]$.

CHAPITRE II

MÉTHODES D'ESTIMATION DE LA PRODUCTION

Supposons que l'on s'intéresse à la production d'une cohorte déterminée d'une certaine espèce hétérotrophe. Pour mesurer la biomasse produite par cette cohorte au cours d'une période $[t_D, t_F]$ donnée, admettons que l'on dispose à la fois de

$$\begin{cases} N_i & = \text{nombre d'individus vivants au moment } t_i \text{ et} \\ m_i & = \text{masse individuelle moyenne de ces sujets} \end{cases}$$

en des instants $t_D = t_0 < t_1 < \dots < t_n < t_{n+1} = t_F$.

L'objectif de ce chapitre est de présenter et d'illustrer quatre méthodes dites directes d'estimation de la production $P_{[t_D, t_F]}$ basées sur un tel échantillon, à savoir la méthode de la sommation des accroissements (section 2.1), la méthode de la sommation des pertes (section 2.2), la méthode de la courbe de Allen (section 2.3), et la méthode du taux de croissance instantanée (section 2.4). Enfin, une cinquième technique s'appuyant sur l'analyse des fréquences de tailles est discutée brièvement à la section 2.5.

Comme on l'a vu au chapitre précédent, la valeur théorique (inconnue) du paramètre $P_{[t_D, t_F]}$ est donnée par l'aire sous la courbe (m_t, N_t) entre les points $t = t_D$ et $t = t_F$ lorsque les fonctions m_t et N_t sont supposées continues et

respectivement monotone croissante et décroissante. Les quatre “méthodes directes” présentées ici partent de ces hypothèses.

2.1 Méthode de la sommation des accroissements

Cette méthode, la plus simple de toutes, consiste à supposer en première approximation que m_t et N_t sont en relation linéaire sur chacun des intervalles $[t_i, t_{i+1}]$, $i = 0, \dots, n$. Si tel était le cas, la production P_i associée à cet intervalle serait tout simplement l'aire du trapèze $m_i m_{i+1} N_i N_{i+1}$, tel qu'illustré à la figure 2.

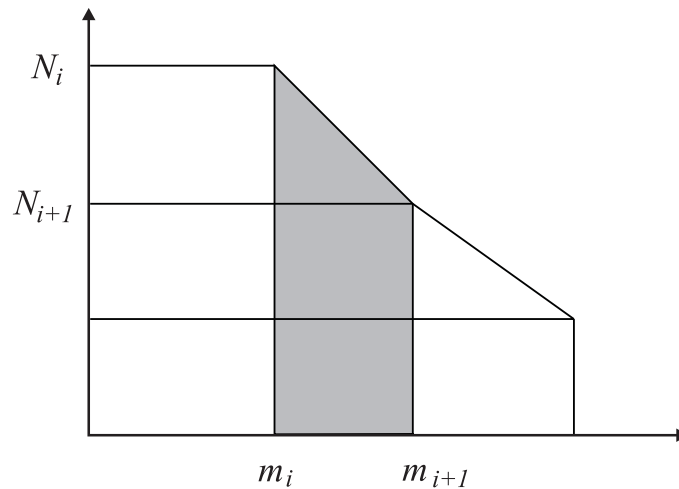


Figure 2 : Surface entre m_i et m_{i+1} . Estimation de la production dans un intervalle de temps $[t_i, t_{i+1}]$ sous l'hypothèse que la masse individuelle moyenne m_t est une fonction linéaire décroissante de N_t sur cet intervalle.

Sous cette hypothèse, une estimation de P_i est donc donnée par

$$\hat{P}_i = \frac{(N_{i+1} + N_i)(m_{i+1} - m_i)}{2},$$

ce qui conduit à proposer

$$\hat{P}_{[t_D, t_F]} = \sum_{i=0}^n \hat{P}_i \quad (2.1)$$

comme estimation de la production sur l'ensemble de l'intervalle $[t_D, t_F]$.

2.2 Méthode de la sommation des pertes

Cette méthode d'estimation, attribuée au Danois Boysen-Jensen (1919), est basée sur l'idée que tout ce qui est produit finit par disparaître, soit par décès, soit par prédation. Ainsi, une estimation de la mortalité (au sens général) devrait conduire à une estimation de la production totale d'une cohorte.

Pour tout $i = 0, \dots, n$, Boysen-Jensen propose d'estimer la biomasse M_i liée à la mortalité intervenue entre les instants t_i et t_{i+1} par l'aire ombrée de la figure 3, à savoir celle du trapèze $m_i m_{i+1} N_i N_{i+1}$. En d'autres termes, il propose une estimation de M_i par interpolation linéaire, soit

$$\hat{M}_i = \frac{(m_{i+1} + m_i)(N_i - N_{i+1})}{2}.$$

Dans ce contexte, il est naturel d'estimer la production totale sur l'intervalle $[t_D, t_F]$ par

$$\tilde{P}_{[t_D, t_F]} = (B_F - B_D) + \sum_{i=0}^n \hat{M}_i,$$

puisque $B_F - B_D$ représente la différence entre la biomasse initiale B_D et la biomasse $B_F = B_{n+1}$ à la dernière date d'échantillonnage. Si t_F est suffisamment grand pour garantir que tous les membres de la cohorte sont éteints, alors $B_F = 0$, mais ceci se produit rarement dans la pratique.

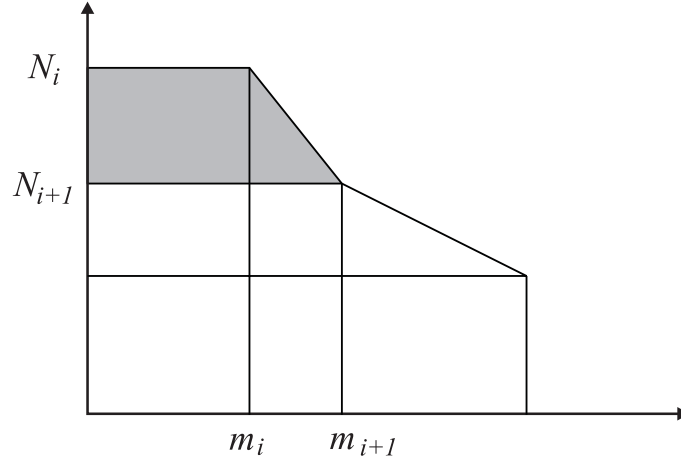


Figure 3 : Surface entre N_i et N_{i+1} . Estimation de la mortalité dans un intervalle de temps $[t_i, t_{i+1}]$ sous l'hypothèse que la masse individuelle moyenne m_t est une fonction linéaire décroissante de N_t sur cet intervalle.

Remarque 1. Les méthodes de la sommation des accroissements et des pertes sont parfaitement identiques. En effet, on voit immédiatement par substitution que

$$\begin{aligned}
 \hat{P}_{[t_D, t_F]} - \sum_{i=0}^n \hat{M}_i &= \sum_{i=0}^n (\hat{P}_i - \hat{M}_i) \\
 &= \frac{1}{2} \sum_{i=0}^n \left\{ (N_{i+1} + N_i)(m_{i+1} - m_i) - \right. \\
 &\quad \left. (N_i - N_{i+1})(m_{i+1} + m_i) \right\} \\
 &= \sum_{i=0}^n N_{i+1} m_{i+1} - \sum_{i=0}^n N_i m_i \\
 &= N_{n+1} m_{n+1} - N_0 m_0 \\
 &= B_F - B_D.
 \end{aligned}$$

Exemple 1. L'exemple suivant (voir tableau 1) de calcul de la production par la méthode de la sommation des accroissements et celle de la sommation des pertes est tiré de l'article de Crisp (1984). Posons :

M = masse moyenne individuelle sèche exprimée en mg,

N = densité de la population exprimée en milliers par mètre carré,

Nm = biomasse exprimée en kilogramme par mètre carré,

$$m^* = \frac{m_i + m_{i+1}}{2}, \quad N^* = \frac{N_i + N_{i+1}}{2}, \quad \Delta N = N_i - N_{i-1},$$

$$\Delta m = m_{i+1} - m_i.$$

On constate à travers les calculs présentés au tableau 1 que les deux méthodes coïncident bel et bien, la production totale étant de 11.7 kg/m² dans les deux cas (à une erreur d'arrondi près).

Date	m	N	N_m	N^*	m^*	ΔN	Δm
1969							
1 avr.	0.2	100	0.02	—	—	—	—
1 mai	5	60	0.30	80	2.6	40	4.8
1 juin	30	24	0.72	42	17.5	36	25
1 juil.	90	20	1.80	22	60	4	60
1 août	200	18	3.60	19	145	2	110
1 sep	350	16	5.60	17	275	2	150
1 oct.	480	15	7.20	15.5	415	1	130
1 nov.	600	14	8.40	14.5	540	1	120
1 déc.	590	13	7.66	13.5	595	1	-10
1970							
1 janv.	580	12	6.96	12.5	585	1	-10
1 mars	610	10	6.10	11	595	2	30
1 mai	630	8	5.04	9	620	2	20
1 juil.	660	3	1.98	5.5	645	5	30
1 nov.	700	2	1.40	2.5	680	1	40
1971							
1 mai	710	1	0.71	1.5	705	1.0	10
1 nov.	720	0.5	0.36	0.75	715	0.5	10
1972							
1 mai	730	0.2	0.15	0.35	725	0.3	10
1 nov.	740	0.1	0.07	0.15	735	0.1	15
1973							
1 mai	755	0	0	0.05	747	0.1	15

Tableau 1a : Données de Crisp. *Données de Crips (1984) et calculs afférents à l'estimation de la production et de la mortalité pendant la période concernée.*

Date	$\Delta P =$ $N^* \Delta m$	$\Delta M =$ $m^* \Delta N$	$\sum \Delta P$	$\sum \Delta M$	$\sum \Delta M$ $+mN$
1969					
1 avr.	—	—	—	—	—
1 mai	0.38	0.10	0.38	0.10	0.40
1 juin	1.05	0.63	1.43	0.73	1.45
1 juil.	1.32	0.24	2.75	0.97	2.77
1 août	2.09	0.29	4.84	1.26	4.86
1 sep	2.55	0.55	7.39	1.81	7.41
1 oct.	2.02	0.42	9.41	2.23	9.43
1 nov.	1.73	0.54	11.14	2.77	11.17
1 déc.	-0.14	0.60	11.00	3.37	11.03
1970					
1 janv.	-0.12	0.58	10.88	3.95	10.91
1 mars	0.33	1.19	11.21	5.14	11.24
1 mai	0.18	1.24	11.39	6.38	11.42
1 juil.	0.17	3.23	11.56	9.61	11.59
1 nov.	0.10	0.68	11.66	10.29	11.69
1971					
1 mai	0.02	0.71	11.70	11.00	11.71
1 nov.	Negl.	0.35	11.70	11.35	11.71
1972					
1 mai	Negl.	0.22	11.70	11.57	11.72
1 nov.	Negl.	0.07	11.70	11.64	11.71
1973					
1 mai	Negl.	0.07	11.70	11.71	11.71

Tableau 1b : Données de Crisp (suite). *Données de Crisp (1984) et calculs afférents à l'estimation de la production et de la mortalité pendant la période concernée.*

2.3 Méthode de la courbe de Allen

Plutôt que de supposer que m_t et N_t sont fonctions linéaires l'une de l'autre, on peut tenter d'inférer la forme de la relation $N_t = f(m_t)$ à partir des $n + 2$ paires de données recueillies. Autrement dit, on cherche d'abord à identifier une fonction \hat{f} qui décrive bien le nuage de points $(m_0, N_0), \dots, (m_{n+1}, N_{n+1})$. La production sur l'intervalle $[t_D, t_F]$ est alors donnée par

$$\hat{P}_{[t_D, t_F]} = \int_{m_0}^{m_{n+1}} \hat{f}(m) dm.$$

Cette façon générale de procéder est appelée *méthode de la courbe de Allen*. Pour l'illustrer, supposons à titre d'exemple que l'on soupçonne une relation linéaire entre les $\sqrt{N_i}$ et les $\log(m_i)$, $i = 0, \dots, n + 1$. Comme en régression linéaire classique, on pourrait alors chercher les constantes a et b qui minimisent la somme de carrés

$$\sum_{i=0}^{n+1} \left\{ \sqrt{N_i} - a - b \log(m_i) \right\}^2.$$

Si on note

$$\overline{\log(m)} = \frac{1}{n+2} \sum_{i=0}^{n+1} \log(m_i) \quad \text{et} \quad \overline{\sqrt{N}} = \frac{1}{n+2} \sum_{i=0}^{n+1} \sqrt{N_i},$$

il est bien connu que les valeurs optimales de a et de b sont alors données par

$$\hat{a} = \overline{\sqrt{N}} - \overline{b \log(m)} \quad \text{et} \quad \hat{b} = \frac{\sum_{i=0}^{n+1} \left\{ \log(m_i) - \overline{\log(m)} \right\} \left(\sqrt{N_i} - \overline{\sqrt{N}} \right)}{\sum_{i=0}^{n+1} \left\{ \log(m_i) - \overline{\log(m)} \right\}^2}.$$

Ainsi, l'estimation correspondante de la production serait égale à

$$\hat{P}_{[t_D, t_F]} = \int_{m_0}^{m_{n+1}} \left\{ \hat{a} + \hat{b} \log(m) \right\}^2 dm. \quad (2.2)$$

Comme la méthode de Allen nécessite le choix d'un modèle pour f et ne conduit que rarement à une estimation explicite de la production, les biologistes ont tendance à lui préférer la méthode du taux de croissance instantanée, qui fait l'objet de la prochaine section.

2.4 Méthode du taux de croissance instantanée

Plutôt que de supposer une relation linéaire entre N_t et m_t sur chaque intervalle $[t_i, t_{i+1}]$, Ricker (1946) et Allen (1949) proposent de modéliser chacune de ces deux quantités en fonction du temps et d'employer la relation (1.3), à savoir

$$\frac{d}{dt} P_t = N_t \frac{d}{dt} m_t,$$

pour déduire une estimation de la production $P_{[t_i, t_{i+1}]}$. Plus spécifiquement, ces auteurs suggèrent de prendre

- a) $m_t = m_{t_i} e^{g_i(t-t_i)}$ pour tout $t \in [t_i, t_{i+1}]$, où $g_i > 0$ est le taux de croissance en masse sur cet intervalle ;
- b) $N_t = N_{t_i} e^{-a_i(t-t_i)}$ pour tout $t \in [t_i, t_{i+1}]$, où $a_i > 0$ est le taux de mortalité sur ce même intervalle.

Il découle immédiatement de a) que $m'_t = g_i m_t$, et donc

$$P'_t = g_i N_t m_t = g_i N_{t_i} m_{t_i} e^{(g_i - a_i)(t-t_i)}$$

pour tout $t \in [t_i, t_{i+1}]$. En intégrant par rapport à t , on trouve alors

$$\begin{aligned} P_{[t_i, t_{i+1}]} &= \int_{t_i}^{t_{i+1}} P'_t dt = \frac{g_i}{g_i - a_i} N_{t_i} m_{t_i} (e^{(g_i - a_i)(t_{i+1} - t_i)} - 1) \\ &= \frac{g_i}{g_i - a_i} (N_{t_{i+1}} m_{t_{i+1}} - N_{t_i} m_{t_i}). \end{aligned} \quad (2.3)$$

Des estimations naturelles de N_{t_i} , $N_{t_{i+1}}$, m_{t_i} et $m_{t_{i+1}}$ sont données par N_i , N_{i+1} , m_i et m_{i+1} . De plus, partant de a), on déduit que

$$\frac{m_{t_{i+1}}}{m_{t_i}} = e^{g_i(t_{i+1} - t_i)},$$

d'où il est possible de tirer une estimation de g_i , soit

$$\hat{g}_i = \frac{1}{t_{i+1} - t_i} \log \left(\frac{m_{i+1}}{m_i} \right).$$

De la même façon, il découle de a) et b) que

$$\frac{N_{t_{i+1}} m_{t_{i+1}}}{N_{t_i} m_{t_i}} = e^{(g_i - a_i)(t_{i+1} - t_i)},$$

ce qui permet d'estimer $g_i - a_i$ par

$$\widehat{g_i - a_i} = \frac{1}{t_{i+1} - t_i} \log \left(\frac{N_{i+1} m_{i+1}}{N_i m_i} \right).$$

En substituant ces différentes estimations dans la formule (2.3), on conclut qu'une estimation raisonnable de la production sur l'intervalle $[t_i, t_{i+1}]$ est donnée par

$$\hat{P}_{[t_i, t_{i+1}]} = \log \left(\frac{m_{i+1}}{m_i} \right) \frac{B_{i+1} - B_i}{\log(B_{i+1}/B_i)},$$

où $B_i = N_i m_i$ et $B_{i+1} = N_{i+1} m_{i+1}$ représentent respectivement les estimations de la biomasse aux temps t_i et t_{i+1} . Une estimation globale de la

production sur l'intervalle $[t_D, t_F]$ serait alors fournie par

$$\hat{P}_{[t_D, t_F]} = \sum_{i=0}^n \log \left(\frac{m_{i+1}}{m_i} \right) \frac{B_{i+1} - B_i}{\log(B_{i+1}/B_i)}. \quad (2.4)$$

Remarque 2. Dans la pratique, les biologistes utilisent plutôt l'estimateur

$$\hat{P}_{[t_i, t_{i+1}]} = \frac{B_i + B_{i+1}}{2} \log \left(\frac{m_{i+1}}{m_i} \right).$$

Cette simplification se justifie du fait que

$$\frac{B_{i+1} - B_i}{\log(B_{i+1}/B_i)} \approx \frac{B_i + B_{i+1}}{2} \quad (2.5)$$

lorsque $t_{i+1} \approx t_i$. Pour démontrer ce résultat, notons qu'il découle des hypothèses a) et b) que sur l'intervalle $[t_i, t_{i+1}]$, $B_t = B_{t_i} e^{cx}$ avec $c = g_i - a_i$ et $x = t - t_i$. Or, il se trouve que

$$\frac{e^{cx} - 1}{cx} \approx \frac{e^{cx} + 1}{2} \approx 1 + \frac{cx}{2}$$

pour $x \approx 0$, puisque ces deux fonctions admettent ce même développement de Taylor à l'ordre 1 autour de l'origine. En prenant $x = t_{i+1} - t_i$, on trouve donc

$$\frac{B_{t_{i+1}} - B_{t_i}}{\log(B_{t_{i+1}}/B_{t_i})} = B_{t_i} \frac{e^{cx} - 1}{cx} \approx B_{t_i} \frac{e^{cx} + 1}{2} = \frac{B_{t_i} + B_{t_{i+1}}}{2}. \quad (2.6)$$

En remplaçant B_{t_i} et $B_{t_{i+1}}$ par leur estimation respective, on est conduit à (2.5), comme Riggs (1963) l'a fait valoir.

En répétant l'analyse précédente pour chacun des intervalles $[t_i, t_{i+1}]$, et en tenant compte de la remarque 2, on voit que les hypothèses a) et b) qui sont au cœur de la méthode du taux de croissance instantanée conduisent à l'estimation suivante de la production sur l'ensemble de la période $[t_D, t_F]$:

$$\hat{P}_{[t_D, t_F]} = \sum_{i=0}^n \frac{B_i + B_{i+1}}{2} \log \left(\frac{m_{i+1}}{m_i} \right). \quad (2.7)$$

Remarque 3. Gillespie & Benke (1979) ont noté que si $N_t = N_{t_i} e^{-at}$ et $m_t = m_{t_i} e^{gt}$ sur un petit intervalle $[t_i, t_{i+1}]$, alors les méthodes de calcul de la production de la sommation des accroissements, de la sommation des pertes et du taux de croissance instantané donnent des résultats quasi-identiques. En effet, en procédant de la même manière qu'en (2.6), on déduit aisément que

$$\frac{m_t - m_{t_i}}{\log(m_t/m_{t_i})} \approx \frac{m_t + m_{t_i}}{2}$$

pour tout $t \in [t_i, t_{i+1}]$, et donc que

$$\frac{B_{t_i} + B_{t_{i+1}}}{2} \log \left(\frac{m_{t_{i+1}}}{m_{t_i}} \right) \approx \frac{(B_{t_i} + B_{t_{i+1}}) (m_{t_{i+1}} - m_{t_i})}{m_{t_{i+1}} + m_{t_i}}. \quad (2.8)$$

Or, pour $x \approx 0$, on a aussi

$$e^{(g-a)x} + 1 \approx \frac{1}{2} (e^{gx} + 1) (e^{-ax} + 1),$$

puisque les deux premiers termes du développement en série de Taylor des membres de gauche et de droite sont identiques. Par conséquent, si $x =$

$t_{i+1} - t_i \approx 0$, on déduit que

$$\begin{aligned} \frac{B_{t_{i+1}} + B_{t_i}}{2} &= \frac{B_{t_i}}{2} (e^{(g-a)(t_{i+1}-t_i)} + 1) \\ &\approx \frac{N_{t_i} m_{t_i}}{2} \frac{(e^{g(t_{i+1}-t_i)} + 1)(e^{-a(t_{i+1}-t_i)} + 1)}{2}, \end{aligned}$$

de sorte que

$$B_{t_{i+1}} + B_{t_i} \approx (N_{t_{i+1}} + N_{t_i}) \frac{m_{t_{i+1}} + m_{t_i}}{2}$$

et donc

$$B_i + B_{i+1} \approx (N_i + N_{i+1}) \frac{m_i + m_{i+1}}{2}.$$

En substituant dans (2.8), il vient

$$\frac{B_i + B_{i+1}}{2} \log \left(\frac{m_{i+1}}{m_i} \right) \approx \frac{(N_i + N_{i+1})(m_{i+1} - m_i)}{2},$$

ce qu'il fallait démontrer.

Exemple 2. L'exemple suivant est tiré de l'article de Waters & Crawford (1973, p. 291), sur le calcul de la production pour l'espèce *Ephemerella subvaria* échantillonnée entre le 26 septembre et le premier juillet avec

G = taux de croissance instantanée,

B^* = biomasse moyenne = $(B_t + B_{t+1})/2$ en g/m^2 ,

B_t = biomasse à l'instant t en g/m^2 ,

m = masse moyenne individuelle en mg,

$P_{[t,t+1]}$ = production entre t et $t + 1$, en g/m^2 .

Dans le tableau ci-dessous, $G = 0.30$ a été calculé comme une croissance de la masse entre le 6 juin et le 1 juillet, soit 24 jours, où la masse maximale

de 29 mg a été atteinte. En effet,

$$G = \frac{29 - 21.59}{30 - 6} = 0.30,$$

$$B^* = \frac{10 - 3(29 - 21.59) + 0.95}{2} \approx 0.48 \text{ g.}$$

Date	B_t (g/m ²)	m (mg)	G	B^* (g/m ²)	$P_{t,t+1}$ (g/m ²)
26 septembre	1.46	0.23	1.52	3.06	4.65
22 octobre	4.65	1.05	0.44	5.67	2.49
27 novembre	6.69	1.64	0.49	8.76	4.29
27 décembre	10.82	2.67	0.29	11.93	3.46
30 janvier	13.03	3.56	0.58	11.58	6.72
12 mars	10.13	6.38	0.60	6.40	3.84
17 avril	2.67	11.67	0.62	1.81	1.12
6 juin	0.95	21.59	0.30	0.48	0.14
Production totale = $\sum P_{t,t+1}$					26.71

Tableau 2 : Données de Waters et Crawford. *Données concernant la production de Ephemarella subvaria. Source : Waters & Crawford (1973, p. 291).*

2.5 Méthode des fréquences de tailles

Cette approche a été introduite par Hynes (1961) pour estimer le niveau de production secondaire d'assemblages multispécifiques de benthos d'eau douce. Son intérêt réside dans l'obtention d'une estimation de la production

secondaire, même lorsqu'il n'est pas possible d'identifier et de suivre des cohortes à l'intérieur de la population étudiée.

Le principe de la méthode de Hynes consiste à travailler sur une cohorte dite "moyenne," obtenue en faisant la moyenne de l'ensemble des histogrammes des fréquences de tailles observées au cours de l'échantillonnage. On obtient ainsi l'équivalent d'une courbe de Allen et l'estimation de la production est ensuite calculée par la méthode de la sommation des pertes appliquée à chacune des classes de tailles successives. On remarque cependant que le terme de différence entre les biomasses initiale et finale est absent du calcul de la production.

Des corrections ont été apportées à la méthode initiale par Hynes & Coleman (1968), puis par Hamilton (1969), qui a précisé ses modalités d'application, et encore par Benke (1979), qui a introduit un terme de calcul additionnel pour les espèces multivoltines. Finalement, Menzies (1980) a proposé une correction dans le cas où la croissance n'est pas linéaire.

Cette méthode peut être appliquée à des populations dont les cohortes sont distinctes mais, selon Benke (1973), elle ne doit pas être privilégiée par rapport aux méthodes directes.

Les trois hypothèses de base qui sous-tendent la méthode de Hynes sont résumées comme suit par Hamilton (1969) :

a) toutes les espèces du groupe doivent avoir le même voltinisme ;¹

¹Le voltinisme se définit comme le nombre moyen de générations d'une espèce observées au cours d'une période donnée, généralement fixée à un an. Ainsi, le voltinisme de l'espèce humaine pourrait être estimé à 0.05 s'il s'écoulait 20 ans en moyenne avant qu'un humain ne se reproduise.

- b) toutes les espèces du groupe doivent avoir une taille maximale comparable ;
- c) la croissance doit être linéaire.

Pour mettre en œuvre cette méthode, il faut d'abord recueillir des échantillons représentatifs de la population à des instants judicieusement répartis dans le temps. On doit ensuite diviser la population en un nombre k quelconque de classes de tailles et recenser le nombre N_i d'individus dans chacune de ces classes. Enfin, le temps total de développement de la population D_{tot} doit être déterminé en fonction du voltinisme de l'espèce. En général, on prend $D_{\text{tot}} = 1 \text{ an} = 365 \text{ jours}$ pour une espèce univoltine.

Le temps de développement D_i de la i ème classe est donné par $D_i = D_{\text{tot}}/i$ et le taux de croissance, supposé linéaire, est donné par

$$\frac{m_{\text{max},i} - m_{\text{min},i}}{D_i},$$

où $m_{\text{max},i}$ et $m_{\text{min},i}$ sont respectivement les masses maximale et minimale des individus de la i ème classe. De plus, la production elle-même est supposée proportionnelle au nombre N_i d'individus de la classe, de sorte que la production P_i de la i ème classe peut être estimée par :

$$\hat{P}_i = N_i \frac{m_{\text{max},i} - m_{\text{min},i}}{D_i}.$$

La production annuelle totale sera alors donnée par

$$\begin{aligned} \hat{P} = \sum_{i=1}^k \hat{P}_i &= \sum_{i=1}^i N_i \frac{m_{\text{max},i} - m_{\text{min},i}}{D_i} \\ &= k \sum_{i=1}^k N_i \frac{m_{\text{max},i} - m_{\text{min},i}}{D_{\text{tot}}}. \end{aligned}$$

Remarque 4. Benke (1979) suggère de corriger les P_i en fonction de l'*intervalle de production de la cohorte* (IPC) en les multipliant par le coefficient $365/\text{IPC}$. L'IPC représente le laps de temps nécessaire, exprimé en mois ou en jours, pour qu'une cohorte (ou une classe) passe de l'éclosion à la dernière classe de taille, où le développement est complet. Ainsi,

$$\hat{P} = k \frac{365}{\text{IPC}} \sum_{i=1}^k N_i \frac{m_{\max,i} - m_{\min,i}}{D_{\text{tot}}} .$$

Cette correction n'est pas nécessaire si $D_{\text{tot}} = 1 \text{ an} = 365 \text{ jours}$.

Exemple 3. L'exemple suivant est tiré du livre de Rigler & Downing (1984, p. 53). Il concerne l'estimation de la production de *Thermocyclops hyalinus* dans le lac George, en Ouganda. Les périodes de développement sont exprimées en jours, les masses en microgrammes et la production en microgramme par jour.

Classe	N_i	D_i	m_{\min}	m_{\max}	N_i/D_i	$m_{\max} - m_{\min}$	P_i
Oeufs	220	1.5	0.00	0.05	147	0.05	7.4
Nauplii	199	6.0	0.04	0.16	33	0.12	4.0
Copepodites	327	11.0	0.16	0.95	30	0.79	23.7

Tableau 3 : Données de Rigler et Downing. *Données concernant la production de Thermocyclops hyalinus dans le lac George, en Ouganda. Source : Rigler & Downing (1984, p. 53).*

Dans ce cas particulier, la production journalière totale est estimée à

$$7.4 + 4 + 23.7 = 35.1 \mu\text{g}$$

et la production annuelle est de

$$365 \times 35.1 \mu\text{g} = 12\,811.5 \mu\text{g} = 12.8 \text{ mg.}$$

2.6 Résumé des méthodes

Un résumé sur le mode d'application des méthodes abordées dans ce chapitre est fourni par Benke (1973) sous la forme suivante.

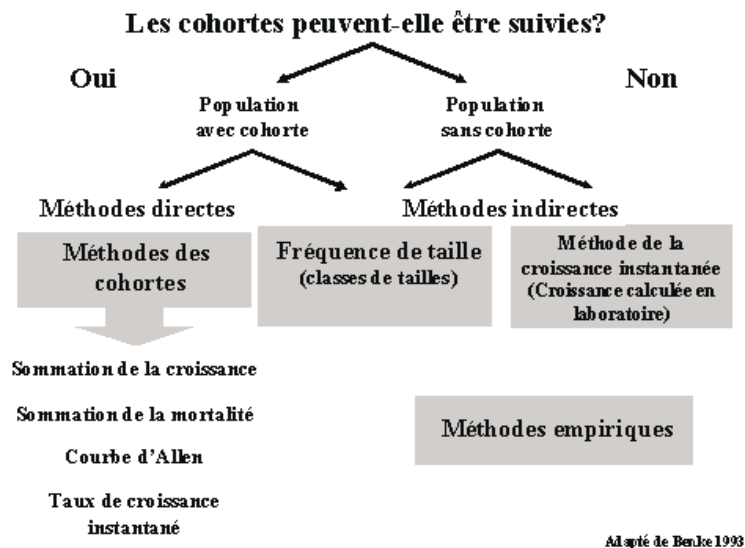


Figure 4 : Logigramme. Logigramme résumant les modalités d'application des principales méthodes d'estimation directes et indirectes de la production secondaire.

CHAPITRE III

QUALITÉ DES ESTIMATEURS: UNE ÉTUDE DE CAS

3.1 Travaux antérieurs

Bien mince est le corpus de connaissances concernant les propriétés statistiques des méthodes directes de production exposées au chapitre 2. Règle générale, les biologistes concernés par les questions de production se contentent de rapporter des estimations brutes. Très peu d'auteurs tentent de mesurer la variabilité de leurs estimations, mais bon nombre d'entre eux prennent quand même la précaution de comparer les réponses auxquelles conduisent les différentes méthodes existantes.

Newman & Martin (1983) rapportent que Chapman (1967) aurait été le premier à proposer un estimateur de variance généralisé pour la méthode du taux de croissance instantanée. Ils mettent eux-mêmes de l'avant des estimateurs de variance pour cette méthode, ainsi que pour celle de la sommation des accroissements, mais pour que les formules puissent être mises en œuvre, il faut procéder par capture-marquage-recapture à chaque date d'échantillonnage.

Faute de mieux, certains auteurs ont tenté de déterminer par simulation, et pour des populations répondant à des caractéristiques bien spécifiques, l'étendue du biais et le degré de variabilité des méthodes directes d'estimation de la production. Vont notamment dans ce sens les travaux de Cushman, Shugart, Hildebrand & Elwood (1977), ou ceux plus récents de Morin, Mousseau & Roff (1987).

Le présent chapitre poursuit dans la même voie en décrivant les résultats d'une petite étude de simulation réalisée sur une seule cohorte de moules bleues (*Mytilus edulis*), dont les caractéristiques sont précisées à la section 3.2. L'expérience effectuée sur ordinateur est ensuite détaillée à la section 3.3, et les résultats sont rapportés à la section 3.4. Nos conclusions sont consignées à la section 3.5.

3.2 Description d'une population fictive

Afin d'évaluer la performance des principales méthodes d'estimation de la production dans un contexte réaliste, un programme informatique réalisé en C++ par M. Jean-François Plante a été utilisé afin de constituer une population théorique comportant au départ 1 000 000 d'individus. Ce programme, conçu sous la direction du Professeur Genest, permet de générer les dimensions ou la masse quotidiennes de chacun des membres d'une population de taille fixée sur une période de temps prédéterminée. Il permet en outre de contrôler la mortalité individuelle et donc de calculer la production de façon exacte.

Sur les conseils du biologiste Mathieu Cusson, chercheur au *Groupe interuniversitaire de recherches océanographiques du Québec* (GIROQ), le programme de M. Plante a été utilisé pour générer une population de moules bleues (*Mytilus edulis*), espèce commune sur les berges du Saint-Laurent dont les caractéristiques biologiques sont bien connues.

Par souci de simplicité, nous avons supposé que tous les individus de la cohorte naissaient le même jour (le premier janvier de l'an 0) et avaient la même taille (mesurée en cm) à la naissance. Selon Bayne & Worrall (1980), la taille moyenne L_t au temps t de la coquille d'une moule de cette espèce peut être approximée de façon satisfaisante par une fonction de Gompertz, définie par

$$\log_{10}(L_t) = a(1 - e^{b(t-c)}), \quad (3.1)$$

où

- a) t représente le temps ;
- b) a représente le logarithme en base 10 de la taille maximale d'une moule de cette espèce ;
- c) b est un taux de croissance et c est un paramètre de translation.

À partir de données recueillies au confluent des estuaires de la Lynher et de la Tamar, près de Plymouth (Angleterre), Bayne & Worrall (1980) ont trouvé $a = 0.838$, $b = -0.584$ et $c = 1.77$ et concluent que ces valeurs sont adéquates, lorsque le temps est mesuré en années et la taille en centimètres. Ainsi, la taille d'une moule à maturité serait de $10^{0.838} \approx 6.887$ cm.²

²D'après ces mêmes auteurs, un modèle de von Bertalanffy serait préférable lorsque les moules approchent de leur maturité. Ce modèle stipule que $L_t = a^*(1 - e^{b^*(t-c^*)})$.

En se servant de l'équation (3.1), il est facile de déterminer quel est, en moyenne, l'accroissement

$$d_i = L_{(i+1)/365} - L_{i/365}$$

d'une moule du jour i au jour $i+1$. Pour introduire un élément de variabilité, on a supposé que la taille $L_{j(i+1)}$ de la $j^{\text{ième}}$ moule le jour $i+1$ était de la forme

$$L_{j(i+1)} = L_{ji} + d_i + \epsilon_{ji},$$

les ϵ_{ji} étant des variables aléatoires gaussiennes indépendantes de moyenne nulle et de variance σ^2 . On a supposé $\sigma = 0.001$ dans les simulations.

Pour convertir la taille quotidienne L_{ji} de la moule en une masse sèche M_{ji} sans coquille, masse exprimée en milligrammes (mg), on a ensuite utilisé la relation

$$M_{ji} = 0.0655L_{ji}^{4.157},$$

tel que recommandé par Bayne & Worrall (1980) pour leur population.

En ce qui a trait à la mortalité, on a supposé que la durée de vie de chaque moule était une variable aléatoire exponentielle d'espérance $10/7 \approx 1.43$ années. De plus, les moules ont été assignées aléatoirement à l'intérieur de 500 quadrats, espace rectangulaire correspondant à une unité d'échantillonnage. Chaque quadrat comportait donc en moyenne 2000 moules.

Au bout du compte, nous avons donc à notre disposition une base de données d'un million de lignes et de $4 \times 365 + 1 = 1461$ colonnes donnant, pour une période de quatre ans, la masse quotidienne de chacune des moules. Deux colonnes supplémentaires indiquaient le jour du décès de chaque moule et le quadrat auquel elle appartenait.

Les figures 5 et 6 illustrent respectivement la mortalité observée et la masse moyenne des vivants en fonction du temps. L'axe du temps est gradué en jours ; l'origine $t = 0$ correspond au premier janvier de l'an 1, c'est-à-dire au moment où les individus célèbrent leur premier anniversaire de naissance. L'évolution des variables N_t et m_t est montrée jusqu'au temps $t = 1095$, qui correspond au premier janvier de l'an 4.

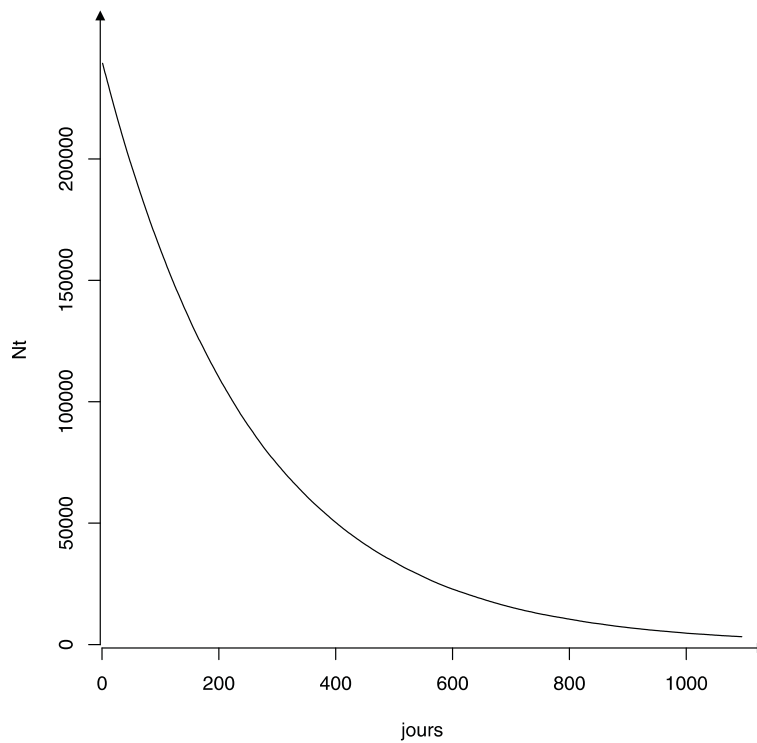


Figure 5 : Évolution de N_t en fonction du temps. Courbe illustrant le nombre de moules vivantes en fonction du temps dans la population simulée ; l'axe du temps est gradué en jours et l'origine $t = 0$ correspond au premier janvier de l'an 1, c'est-à-dire au moment où les individus célèbrent leur premier anniversaire de naissance.

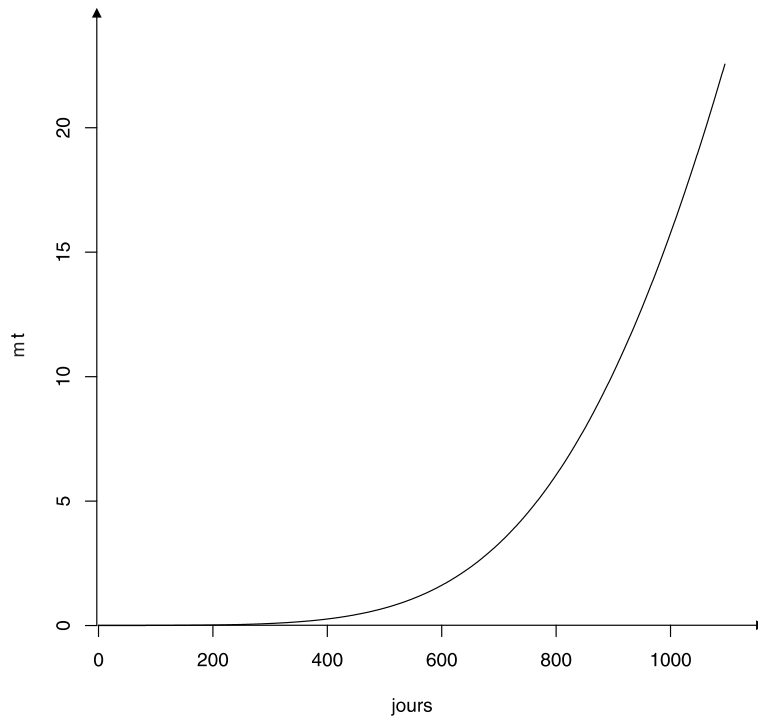


Figure 6 : Évolution de m_t en fonction du temps. *Courbe illustrant l'évolution de la masse moyenne individuelle dans la population simulée ; l'axe du temps est gradué en jours et l'origine $t = 0$ correspond au premier janvier de l'an 1, c'est-à-dire au moment où les individus célèbrent leur premier anniversaire de naissance.*

Les nombres initial et final de moules pendant la période retenue pour la cueillette de données sont de

$$N_0 = 239\,390 \quad \text{et} \quad N_{1095} = 3238,$$

tandis que leur masse moyenne passe de

$$m_0 = 7.18 \times 10^{-4} \text{ mg} \quad \text{à} \quad m_{1095} = 22.56 \text{ mg.}$$

L'aspect lisse des courbes expérimentales tracées aux figures 5 et 6 s'explique par la très grande population de vivants sur laquelle elles s'appuient. Noter toutefois qu'en raison de la variation aléatoire individuelle propre au modèle, les masses moyennes observées aux temps 0 et 1095 sont différentes des valeurs théoriques, à savoir

$$E(m_0) = 7.03 \times 10^{-4} \text{ mg} \quad \text{et} \quad E(m_{1095}) = 22.60 \text{ mg}.$$

3.3 Description de l'expérience

Il est facile, au moyen de la base de données, de déterminer la production réelle entre le premier janvier de l'an 1 (temps $t_D = 0$) et le premier janvier de l'an 4 (temps $t_F = 1095$). Il suffit pour ce faire de prendre la différence entre la masse finale et la masse initiale de toutes les moules vivantes au premier janvier de l'an 4 ou décédées avant cette date. En sommant sur l'ensemble des individus de la population, on a ainsi obtenu la valeur théorique de la production, à savoir

$$P_{[t_D, t_F]} = 221.45 \text{ g}.$$

Afin de nous faire une idée de l'influence de la fréquence et de l'effort d'échantillonnage sur le biais et la variance des trois méthodes directes d'estimation de la production décrites au chapitre 2, nous avons ensuite considéré 12 scénarios différents.

D'une part, quatre fréquences d'échantillonnage ont été fixées, à savoir

- une fois l'an, le premier janvier,
- deux fois l'an, le premier janvier et le premier juillet,

- trois fois l'an, le premier janvier, le premier mai et le premier septembre,
- quatre fois l'an, le premier janvier, le premier avril, le premier juillet et le premier octobre,

étant entendu qu'il y avait échantillonnage de toute façon les premiers janvier des ans 1 et 4.

D'autre part, trois niveaux d'effort d'échantillonnage ont été envisagés, soit un, cinq ou dix quadrats choisis au hasard par date d'échantillonnage. Dans tous les cas, on en a déduit une seule valeur (moyenne) de l'effectif N_t par quadrat et de la masse moyenne m_t de ces moules.

La cueillette de données a été répétée 1000 fois pour chaque fréquence et chaque effort d'échantillonnage. Pour chaque méthode d'estimation, on en a alors déduit 1000 valeurs estimées du paramètre

$$\theta = \frac{P_{[t_D, t_F]}}{500} = \frac{221.45 \text{ g}}{500 \text{ quadrats}} = 442.91 \text{ mg/quadrat},$$

c'est-à-dire de la production totale moyenne par quadrat. Ces valeurs ont ensuite été portées sur un histogramme, afin d'en évaluer le biais et la variabilité.

Compte tenu du manque de linéarité manifeste de la relation entre les variables N_t et m_t dans la population simulée (voir figure 7), un cinquième estimateur a été considéré. Celui-ci est basé sur la méthode de Allen appliquée, non pas aux données brutes, mais à des valeurs transformées de N_t et de m_t afin d'assurer une relation linéaire entre elles.

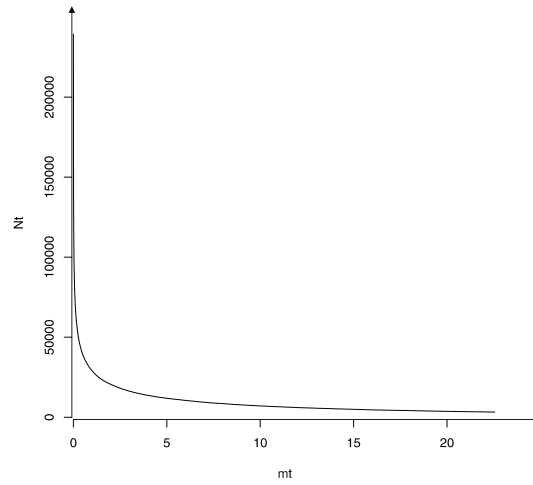


Figure 7 : Relation entre N_t et m_t . Courbe illustrant la relation existant entre N_t et m_t dans la population simulée entre les temps $t_D = 0$ et $t_F = 1095$.

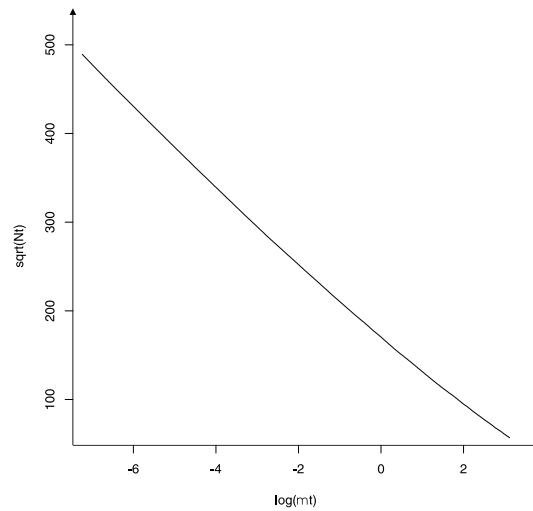


Figure 8 : Relation entre $\sqrt{N_t}$ et $\log(m_t)$. Courbe illustrant la relation existant entre $\sqrt{N_t}$ et $\log(m_t)$ dans la population simulée entre les temps $t_D = 0$ et $t_F = 1095$.

De façon plus spécifique, la figure 8 suggère que $\sqrt{N_t} = \sqrt{f(m_t)}$ peut être approché par une fonction linéaire de $\log(m_t)$. Une estimation raisonnable de la production entre les instants t_D et t_F devrait donc être donnée par

$$\int_{m_0}^{m_{1095}} \hat{f}(m) dm \quad \text{avec} \quad \sqrt{\hat{f}(m)} = a + b \log(m), \quad (3.2)$$

les constantes a et b étant choisies par la méthode des moindres carrés. En d'autres termes, a et b minimisent la somme de carrés des erreurs entre les valeurs observées de $\sqrt{N_i}$ et celles de $a + b \log(m_i)$. Le nombre de valeurs sur lesquelles s'appuient ces estimations dépend de la fréquence avec laquelle la population est échantillonnée, mais pas du nombre de quadrats choisis, puisque N_i et m_i représentent des valeurs moyennes par quadrat.

3.4 Résultats

Les figures suivantes résument les résultats de notre simulation sous forme graphique. Chacun page correspond à l'un des estimateurs considérés, à savoir :

- Figure 9 : la méthode (2.1) de sommation des accroissements ;
- Figure 10 : la méthode (2.2) de la courbe de Allen basée sur la régression linéaire simple entre N_t et m_t ;
- Figure 11 : la méthode (3.2) de la courbe de Allen basée sur la régression linéaire entre $\sqrt{N_t}$ et $\log(m_t)$;
- Figure 12 : la méthode du taux de croissance instantanée, version (2.4) ;
- Figure 13 : la méthode du taux de croissance instantanée, version (2.7).

Chaque figure est constituée de douze histogrammes, répartis en trois lignes et quatre colonnes lorsque la feuille est placée en format “paysage.” Les lignes correspondent aux efforts d’échantillonnage (1, 5 ou 10 quadrats), tandis que les colonnes font référence à la fréquence de cueillette (1, 2, 3 ou 4 fois l’an).

Les conditions précises d’échantillonnage sont rappelées au bas de chaque dessin, qui illustre la variation de l’estimateur spécifié à la faveur de 1000 répétitions indépendantes. On trouve aussi au bas de chaque histogramme le biais et la variance observés de l’estimateur. Ceux-ci sont respectivement définis par

$$\text{biais} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta) \equiv \bar{\theta} - \theta$$

et

$$\text{variance} = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \bar{\theta})^2,$$

où $\hat{\theta}_1, \dots, \hat{\theta}_{1000}$ représentent les 1000 valeurs estimées du paramètre $\theta = 442.91$ mg/quadrat.

Les valeurs observées du biais et de la variance des cinq estimateurs considérées sont également colligées dans les tableaux 4 et 5, en pp. 43 et 44.

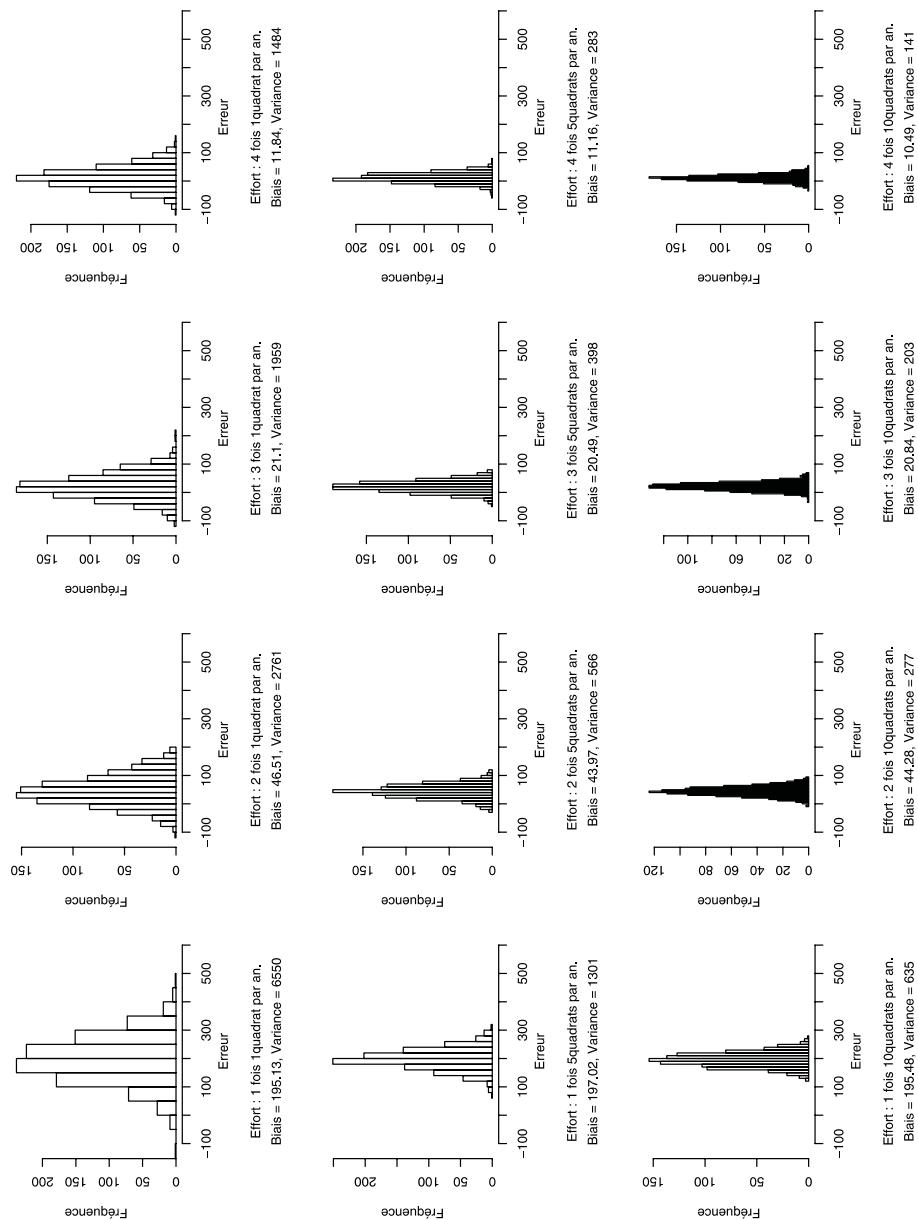


Figure 9 : Résultats pour la méthode de la sommation des accroissements. Histogrammes montrant la variation de l'estimateur (2.1) lors de 1000 répétitions de douze scénarios d'échantillonnage.

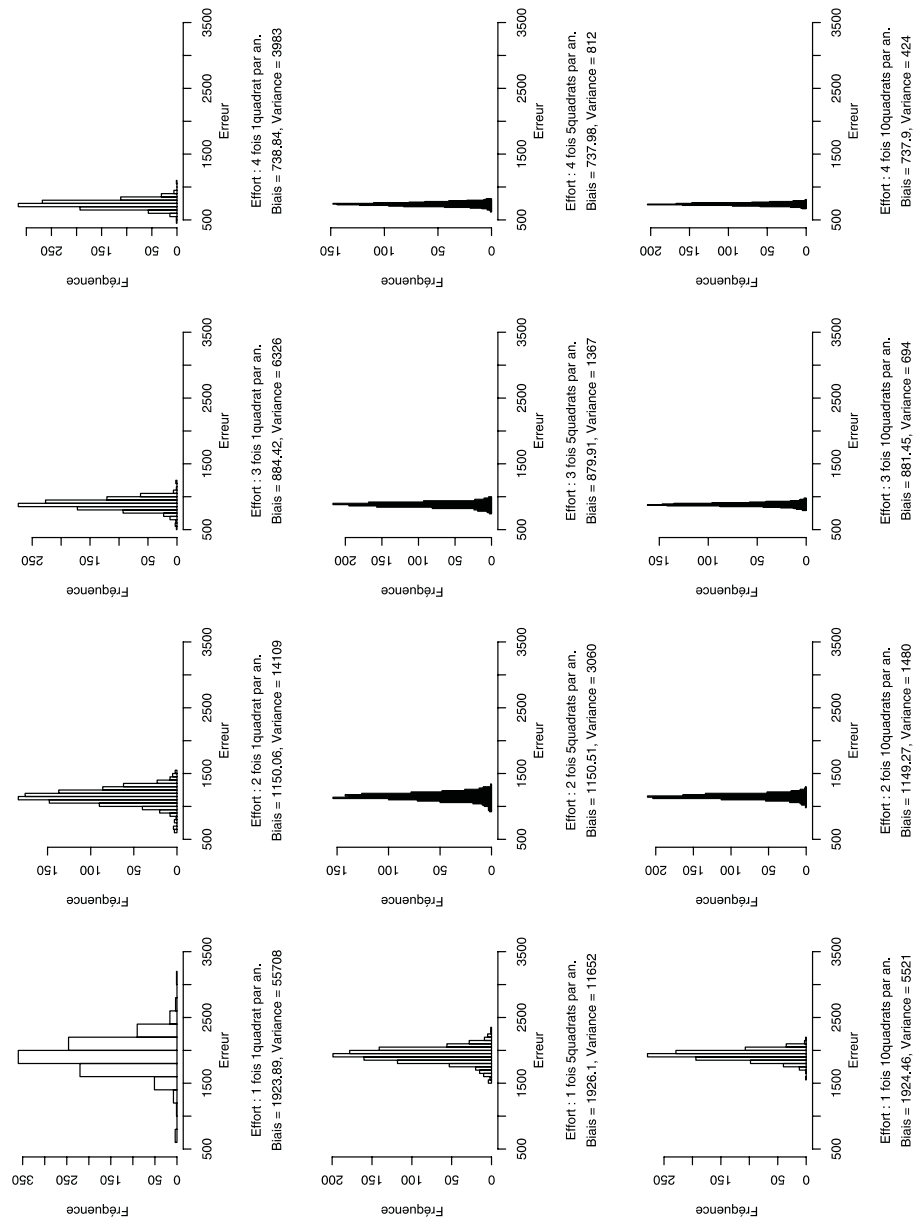


Figure 10 : Résultats pour la méthode de la courbe de Allen (version naïve). Histogrammes montrant la variation de l'estimateur (2.2) lors de 1000 répétitions de douze scénarios d'échantillonnage.

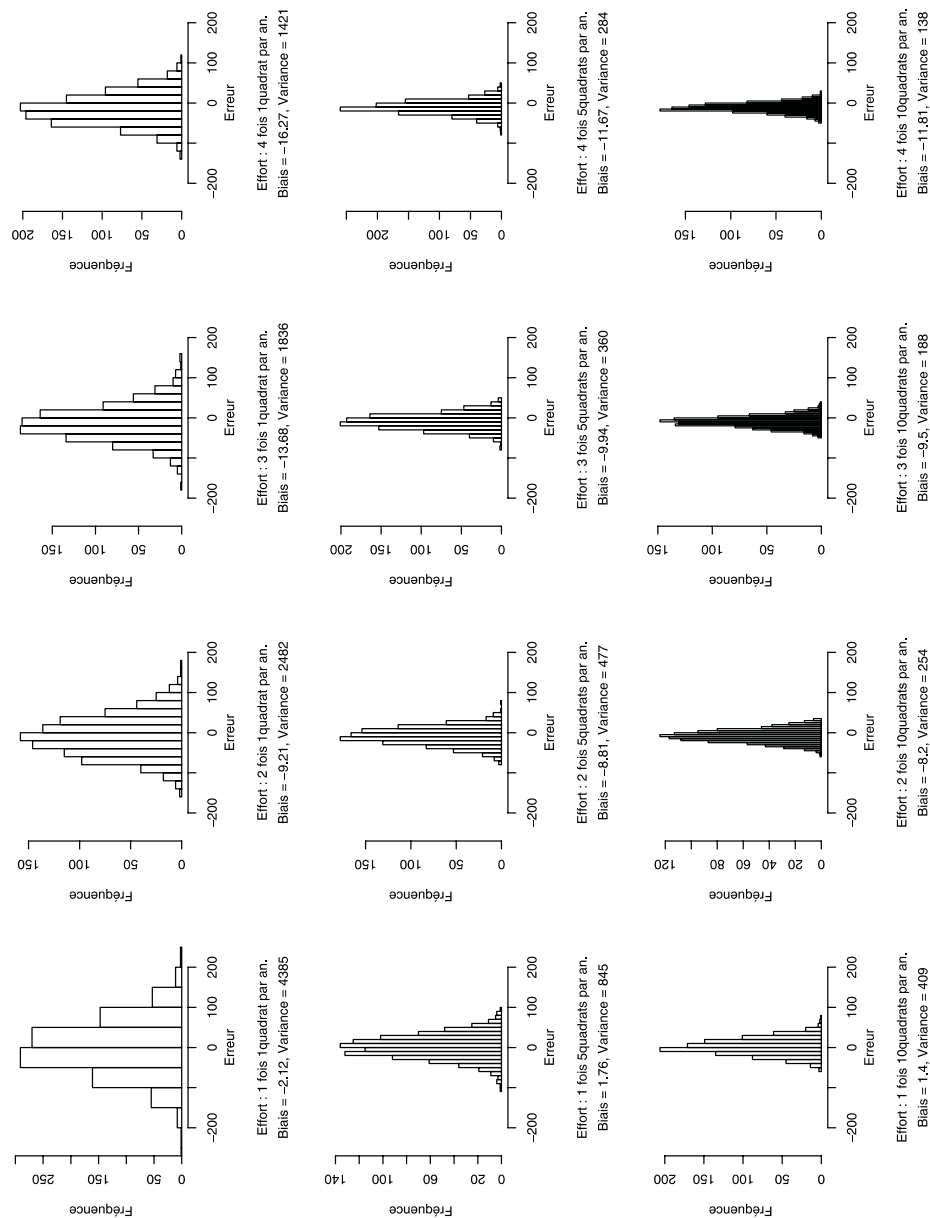


Figure 11 : Résultats pour la méthode de la courbe de Allen (version sophistiquée). Histogrammes montrant la variation de l'estimateur (3.2) lors de 1000 répétitions de douze scénarios d'échantillonnage.

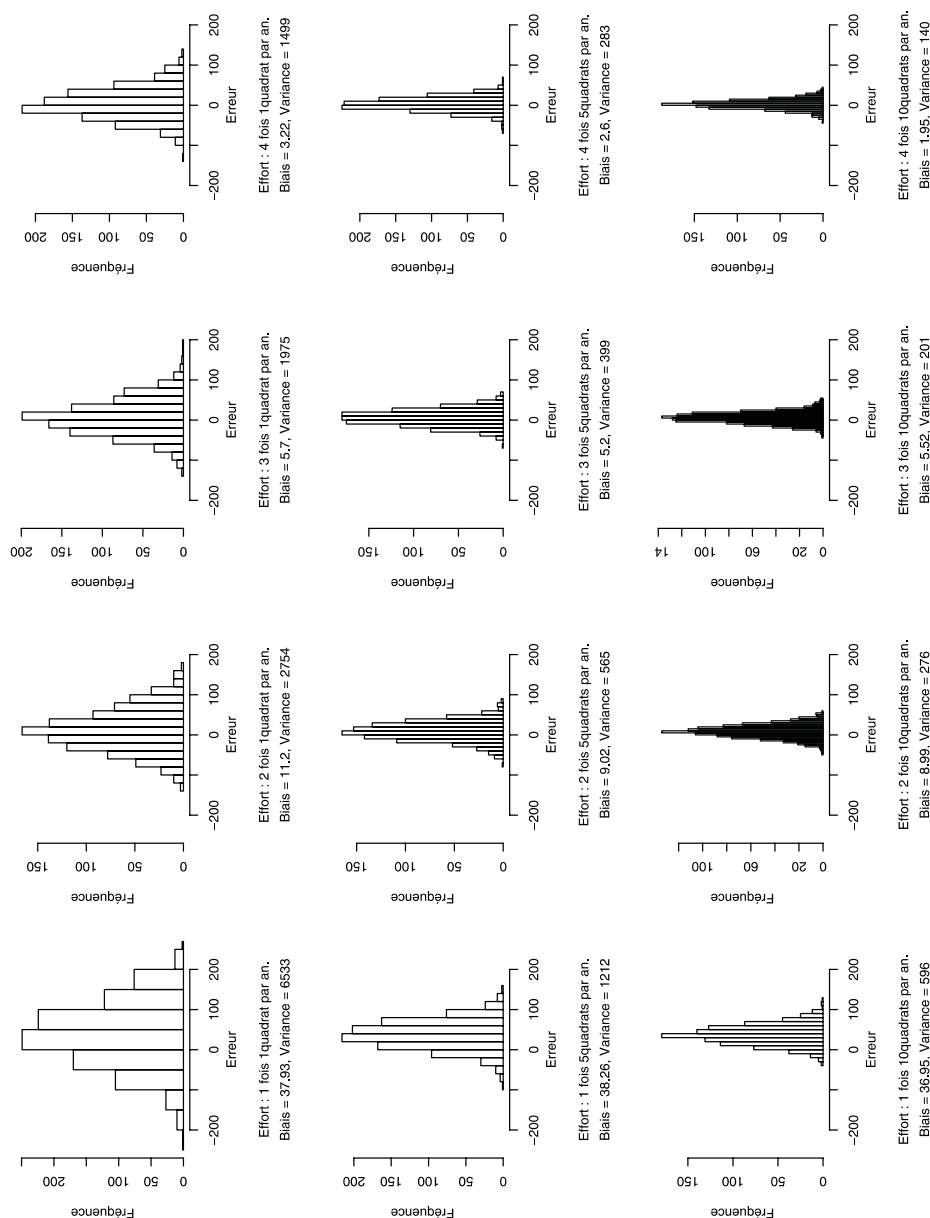


Figure 12 : Résultats pour la méthode du taux de croissance instantanée (sans approximation). Histogrammes montrant la variation de l'estimateur (2.4) lors de 1000 répétitions de douze scénarios d'échantillonnage.

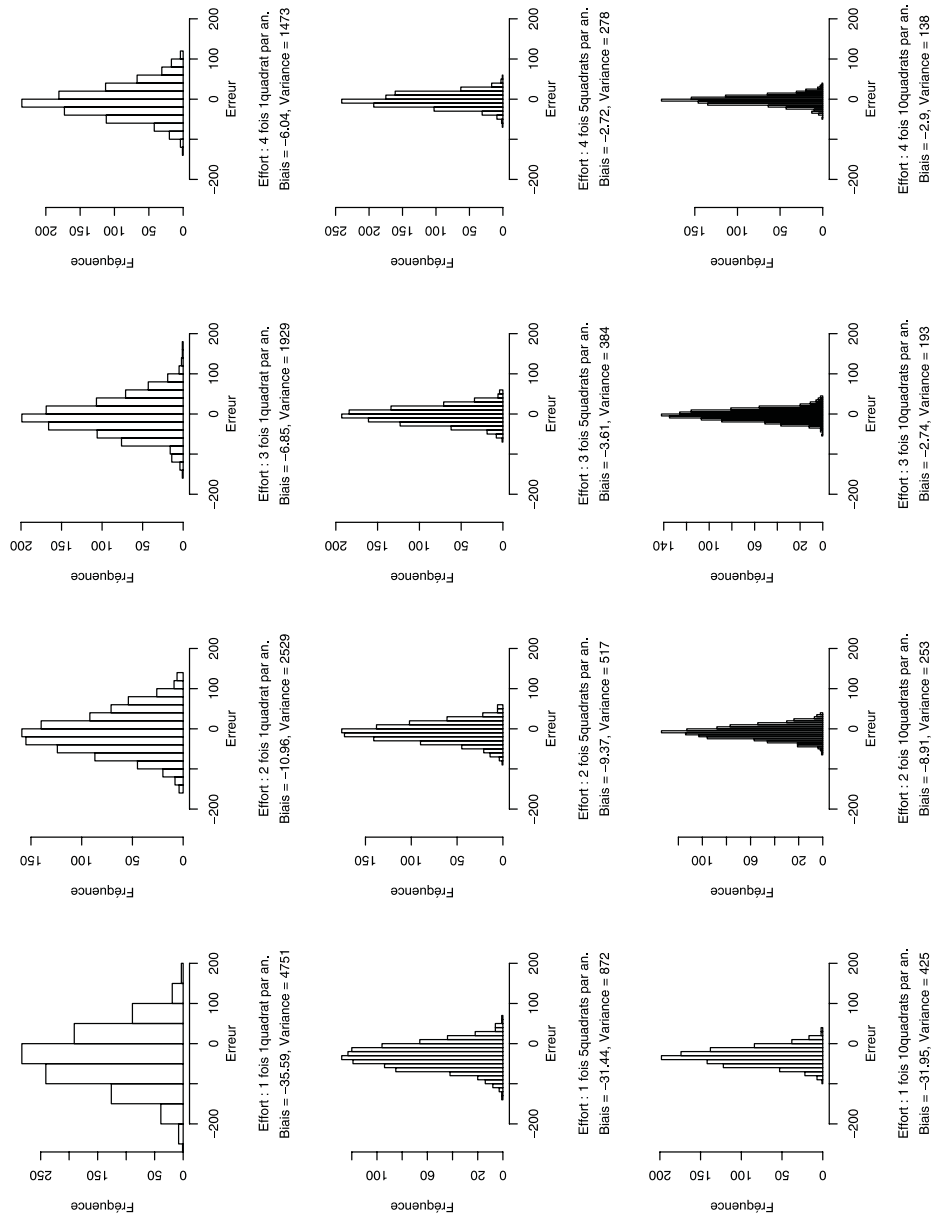


Figure 13 : Résultats pour la méthode du taux de croissance instantanée (avec approximation). Histogrammes montrant la variation de l'estimateur (2.7) lors de 1000 répétitions de douze scénarios d'échantillonnage.

1 quadrat par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	195.13	46.51	21.10	11.84
Allen (2.2)	1923.89	1150.06	884.42	738.84
Allen (3.2)	-2.12	-9.21	-13.68	-16.27
Croissance (2.4)	-35.59	-10.96	-6.85	-6.04
Croissance (2.7)	37.93	11.20	5.70	3.22

5 quadrats par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	197.02	43.97	20.49	11.16
Allen (2.2)	1926.10	1150.51	879.91	737.98
Allen (3.2)	1.76	-8.81	-9.94	-11.67
Croissance (2.4)	-31.44	-9.37	-3.61	-2.72
Croissance (2.7)	38.26	9.02	5.20	2.60

10 quadrats par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	195.48	44.28	20.84	10.49
Allen (2.2)	1924.46	1149.27	881.45	737.90
Allen (3.2)	1.40	-8.20	-9.50	-11.81
Croissance (2.4)	-31.95	-8.91	-2.74	-2.90
Croissance (2.7)	36.95	8.99	5.52	1.95

Tableau 4 : Biais des estimateurs. *Biais observé (en mg) des cinq estimateurs en fonction de la fréquence et de l'effort d'échantillonnage.*

1 quadrat par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	6550	2761	1959	1484
Allen (2.2)	55708	14109	6326	3983
Allen (3.2)	4385	2482	1836	1421
Croissance (2.4)	4751	2529	1929	1473
Croissance (2.7)	6533	2754	1975	1499

5 quadrats par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	1301	566	398	283
Allen (2.2)	11652	3060	1367	812
Allen (3.2)	845	477	360	284
Croissance (2.4)	872	517	384	278
Croissance (2.7)	1212	565	399	283

10 quadrats par date				
Fréquence (fois/an)	1 fois	2 fois	3 fois	4 fois
Accroissements (2.1)	635	277	203	141
Allen (2.2)	5521	1480	694	424
Allen (3.2)	409	254	188	138
Croissance (2.4)	425	253	193	138
Croissance (2.7)	596	276	201	140

Tableau 5 : Variance des estimateurs. *Variance observée (en mg^2) des cinq estimateurs en fonction de la fréquence et de l'effort d'échantillonnage.*

3.5 Discussion

À l'examen des tableaux 4 et 5 ainsi que des figures 9 jusqu'à 13, on peut tirer les conclusions suivantes.

1. *La variabilité de toutes les estimations diminue à mesure que s'accroît la fréquence de l'échantillonnage.* Par exemple, il semble que l'on puisse rabattre la variance de tous les estimateurs par un facteur de deux lorsque l'on augmente la fréquence d'échantillonnage de une à deux fois par an.
2. *La variabilité de toutes les estimations diminue aussi à mesure que s'accroît l'effort d'échantillonnage.* Il semble notamment préférable d'échantillonner cinq quadrats une fois l'an, plutôt qu'un seul quadrat tous les six mois.
3. *La version (3.2) de l'estimateur de Allen constitue le meilleur choix en ce qui concerne la taille du biais d'estimation.* À mesure que s'accroissent l'effort et la fréquence de l'échantillonnage, on constate toutefois que cet estimateur comporte un biais baissier systématique de l'ordre de 13 mg par quadrat sur trois ans dans de grands échantillons. En procédant à un recensement quotidien de la population, on constate en effet que la version (3.2) de la méthode de Allen estime la production totale à 429.05 mg par quadrat pour trois ans, alors que $\theta = 442.91$ dans la population, soit un écart d'environ 3%.

4. *La méthode (2.1) de la sommation des accroissements et les deux versions de celle du taux de croissance instantanée (2.4) et (2.7) sont asymptotiquement sans biais.*
5. *La méthode de Allen naïve (2.2), basée sur la régression linéaire simple entre les valeurs observées N_i et m_i , est de loin la pire.* En effet, elle comporte un biais haussier important, évalué à $744.33 - 442.91 = 301.42$ mg par quadrat sur trois ans, sur la base d'un recensement quotidien de la population.

En résumé, il semble que pour la population considérée, l'estimateur à privilégier soit la version (3.2) de la méthode de Allen, du moins si l'on dispose de ressources d'échantillonnage limitées. La présence d'un biais asymptotique dans cette méthode s'explique sans doute par l'existence d'une très légère (mais réelle) courbure dans la relation entre $\sqrt{N_t}$ et $\log(m_t)$, tel qu'illustré à la figure 8.

CONCLUSION

L'objectif de cet essai était double. D'une part, nous avons voulu y présenter la notion de production et la façon dont les biologistes s'y prennent pour l'estimer par des méthodes directes ou indirectes fondées sur l'échantillonnage. D'autre part, nous avons fait état de nouveaux résultats de simulation permettant d'évaluer le biais et la variance des méthodes directes d'estimation en fonction de l'effort et de la fréquence d'échantillonnage, en prenant comme appui une population fictive dont les caractéristiques de croissance et de mortalité s'apparentaient à celles de la moule bleue (*Mytilus edulis*).

Il est évident que les conclusions que nous avons tirées de notre étude sont très étroitement liées aux hypothèses de base que nous avons émises, concernant notamment le choix de :

- a) la forme et les paramètres des courbes de croissance et de mortalité de la population ;
- b) la valeur de l'écart-type ($\sigma = 0.001$) régissant la variation aléatoire dans l'accroissement quotidien de la taille des sujets ;
- c) se limiter à une seule cohorte et à une même date de naissance pour l'ensemble des individus qui la composent.

Il serait certainement fort instructif de prolonger ces travaux en vérifiant la robustesse de nos conclusions aux changements de lois et de paramètres

pour la croissance et la mortalité, et en quantifiant l'effet d'une augmentation de l'hétérogénéité de la population contrôlée par le paramètre σ .

Par ailleurs, des scénarios à plusieurs cohortes ou à natalité étalée dans le temps pourraient être envisagés, à condition toutefois de ne plus se limiter aux méthodes directes d'estimation de la production.

BIBLIOGRAPHIE

- K. R. Allen (1949). Some aspects of the production and cropping of fresh waters. *Transactions of the Royal Society of New Zealand*, 77, 222–228.
- B. L. Bayne & C. M. Worrall (1980). Growth and production of mussels *Mytilus edulis* from two populations. *Marine Ecology Progress Series*, 3, 317–328.
- A. C. Benke (1973). Concepts and patterns of invertebrate production in running waters. *Verhandlungen der Internationalen Vereinigungen für Limnologie*, 25, 15–38.
- A. C. Benke (1979). A modification of the Hynes method for estimating secondary production with particular significance for multivoltine population. *Limnology and Oceanography*, 24, 168–171.
- P. Boysen-Jensen (1919). Valuation of the Limfjord. 1. *Reports of the Danish Biological Station*, 26, 3–44.
- D. W. Chapman (1967). Production in fish populations. Dans *The Biological Basis of Freshwater Fish Production*, publié sous la direction de S. D. Gerking. Blackwell, Oxford, Angleterre, pp. 3–29.
- D. J. Crisp (1984). Energy flow measurements. Dans *Methods for the Study of Marine Benthos*, deuxième édition publiée sous la direction de N. A.

- Holme et de A. D. McIntyre. Blackwell, Oxford, Angleterre, pp. 284–372.
- R. M. Cushman, H. H. Shugart, S. G. Hildebrand & J. W. Elwood (1977). The effect of growth curve and sampling regime on instantaneous-growth, removal-summation, and Hynes/Hamilton estimates of aquatic insect production: A computer simulation. *Limnology and Oceanography*, 23, 185–189.
- D. M. Gillespie & A. C. Benke (1979). Method for calculating cohort production from field data — some relationships. *Limnology and Oceanography*, 24, 171–176.
- A. L. Hamilton (1969). On estimating annual production. *Limnology and Oceanography*, 14, 771–782.
- H. B. N. Hynes (1961). The invertebrate fauna of a Welsh mountain stream. *Archiv für Hydrobiologia*, 57, 344–388.
- H. B. N. Hynes & M. J. Coleman (1968). A simple method of assessing the annual production of stream benthos. *Limnology and Oceanography*, 13, 569–573.
- C. A. Menzie (1980). A note on the Hynes method of estimating secondary production. *Limnology and Oceanography*, 25, 770–773.
- A. Morin, T. A. Mousseau & D. A. Roff (1987). Accuracy and precision of secondary production estimates. *Limnology and Oceanography*, 32, 1342–1352.

- R. M. Newman & F. B. Martin (1983). Estimation of fish production rates and associated variances. *Canadian Journal of Fisheries and Aquatic Sciences*, 40, 1729–1736.
- W. E. Ricker (1946). Production and utilization of fish population. *Ecological Monographs*, 16, 373–391.
- D. S. Riggs (1963). *The Mathematical Approach to Physiological Problems*. MIT Press, Cambridge, MA.
- F. H. Rigler & J. A. Downing (1984). The calculation of secondary production. Dans *A Manual on Methods for the Assessment of Secondary Productivity in Fresh Waters*. Blackwell, Oxford, Angleterre, pp. 19–86.
- T. F. Waters & G. W. Crawford (1973). Annual production of a stream mayfly population: A comparison of methods. *Limnology and Oceanography*, 18, 286–296.