

**Project title:** Improve Scientific and Technical Advices for fisheries Management (ISTAM)

Proposal/Contract no.: 022774

**Work Package 3:** Stock Assessment Methods and Analysis Tools

**Deliverable D. 3.2.** Standard procedures for stock assessment and case studies

## **Scientific report**

<h3><b>Conversion taille/âge par décomposition polymodale des fréquences de taille</b></h3>
---

**November 2008**

**Chassot E <sup>(1,2)</sup>, Rouyer T <sup>(2,4)</sup>, Sidibé A <sup>(3)</sup>, Gascuel D <sup>(2)</sup>**

<sup>(1)</sup> Institut de Recherche pour le Développement, CRH, Avenue Jean Monnet, BP 171, 34 203 Sète cedex, FRANCE

<sup>(2)</sup> Agrocampus OUEST, UMR 985 INRA-Agrocampus OUEST Ecologie et Santé des Ecosystèmes, 65 rue de St-Brieuc, CS 84 215, 35 042 Rennes cedex, FRANCE

<sup>(3)</sup> Commission Sous-Régionale des Pêches, Amitié 3, Villa 4430, BP 25485, Dakar, SENEGAL

<sup>(4)</sup> Ifremer, CRH, Avenue Jean Monnet, BP 171, 34 203 Sète cedex, FRANCE

# 1. Introduction

La disponibilité des données de captures par groupe d'âge constitue un préalable indispensable à toute évaluation de stock s'appuyant sur une approche structurale. L'estimation des captures aux âges repose sur la connaissance des fréquences de taille au sein des captures qui sont elles-mêmes estimées par échantillonnage. Parmi les méthodes de conversion des structures de tailles en structure d'âges, l'une des plus anciennes et des plus courantes est l'utilisation de clés taille/âge qui indiquent la proportion de chacun des groupes d'âges au sein de chacune des classes de taille (Fridriksson 1934). Ces clés sont établies à partir d'une détermination de l'âge, soit par décomposition polymodale des fréquences de taille, soit par lecture de pièces anatomiques dures. Dans le cas des pêcheries d'Afrique de l'Ouest, un certain nombre de problèmes de conversion des captures en taille en captures aux âges se pose du fait de situations caractérisées par la faible qualité ou même l'absence de données de lectures directes (e.g. otolithes) et les difficultés d'âgeage pour un certain nombre d'espèces (e.g. sardinelles). Dans ce contexte, les méthodes de décomposition polymodale sous contraintes constituent des outils flexibles pour la conversion taille/âge des données de captures (Gascuel 1994). Elles reposent sur le choix d'une loi de répartition des tailles aux âges, d'un modèle de croissance et d'un modèle de variation des écarts types de taille en fonction de l'âge.

Dans le présent document, nous utilisons le package Mixdist du logiciel de statistiques R en libre distribution (R Development Core Team 2008) pour effectuer des décompositions polymodales. Ce package a été développé par Du (2002) et peut être téléchargé sur le site web <http://www.math.mcmaster.ca/peter/mix/mix.html>. Les objectifs sont ainsi de (1) présenter la librairie 'mixdist', (2) développer la méthode sous le logiciel libre R et (3) l'appliquer au cas d'étude du stock de bobo (*Pseudolithus elongatus*) en Guinée Conakry pour en illustrer l'intérêt dans le cas de certaines pêcheries ouest-africaines. Les scripts développés sous le logiciel R développés sont fournis en annexe. Cette sous-activité est complémentaire de l'activité amont "Extrapolation des fréquences de taille", qui s'intéresse aux problèmes d'extrapolation des captures par classe de taille à l'échelle du stock, et de la sous-activité aval "Pseudo-cohorte rectifiée", afin de tester le sensibilité des diagnostics scientifiques à l'estimation des captures par groupe d'âge.

## 2. La librairie 'mixdist'

### 2.1. Généralités

'Mixdist' ajuste des mélanges ("mixtures") de distributions à des données (Rouyer 2006). De façon simple, il ajuste un modèle composé d'une somme de lois de distributions à un jeu de données. La distribution du jeu de données est prise semblable à un mélange de composantes en nombre fini, qui sont modélisées par autant de fonctions de densité de probabilité. On est souvent confronté à ce type de distributions lorsqu'on échantillonne des populations non-homogènes. Ces modèles sont donc adaptés au cas des structure en taille d'une population de poissons qui sont non-homogènes puisque les individus peuvent appartenir à différentes classes d'âge caractérisées par différentes distributions en taille. Sur l'histogramme des fréquences on interprète alors chaque "pic" comme le mode d'une distribution correspondant à une classe d'âge. On s'intéresse ici à des cas où les distributions ont un nombre fini de composantes.

### 2.2. Ajustement de distributions sous contraintes

Mixdist réalise l'ajustement de mélanges de distributions. Chaque modèle est composé d'un type de distribution choisi parmi les distributions normale, log-normale, de Weibull, binomiale, binomiale négative et de Poisson. Ceci nécessite donc l'estimation des proportions relatives à chaque distribution ainsi que la moyenne et l'écart-type de chaque distribution. Pour des raisons pratiques et théoriques il n'est pas possible d'estimer tous les paramètres, particulièrement lorsque les composantes se chevauchent. On réduit donc le nombre de paramètres à estimer grâce aux contraintes sur les paramètres.

#### . Contrainte sur les moyennes

On considère généralement que la loi de répartition des tailles aux âges est une loi normale. La proportion des poissons appartenant à une classe de taille indiquée  $l$ , parmi ceux du groupe d'âge conventionnel  $t$  est alors donnée par :

$$\varphi = \frac{1}{\sigma(t)\sqrt{2\pi}} \exp\left(\frac{-(L_l - L(t))^2}{2\sigma^2(t)}\right) \quad (1)$$

Où  $L_l$  est la longueur moyenne de la classe de taille (en cm),  $L(t)$  est la longueur moyenne à l'âge  $t$  et  $\sigma(t)$  est l'écart type de taille à l'âge  $t$ .

Généralement, la taille moyenne à l'âge  $t$  est connue par l'intermédiaire d'un modèle de croissance linéaire, le modèle le plus courant étant celui de Von Bertalanffy (1938) :

$$L(t) = L_\infty \left(1 - \exp(-K(t - t_0))\right) \quad (2)$$

Où  $L_\infty$  est la longueur asymptotique (ou longueur moyenne des animaux d'âge théorique infini),  $K$  est le coefficient de vitesse de croissance et  $t_0$  l'âge théorique de longueur nulle.

La connaissance a priori des paramètres du modèle de croissance du stock par ajustement à des données taille/âge ou par d'autres méthodes (voir synthèse de Sidibé 2003) permet ainsi de fixer les valeurs des moyennes des lois de répartition. Notez que la valeur de  $t_0$  peut être calculée afin de "caler" la courbe de croissance pour faire coïncider le premier mode de l'histogramme des fréquences de taille avec la taille moyenne au premier âge d'exploitation suivant :

$$t_0 = t + \frac{\ln\left(1 - \frac{L_t}{L_\infty}\right)}{K} \quad (3)$$

#### . Contrainte sur les écarts types

Un modèle décrivant les variations des écarts type de taille en fonction de l'âge peut être utilisé afin d'ajouter une contrainte supplémentaire pour l'ajustement des modèles de mélange. Un modèle comportant un terme constant, un terme proportionnel à la taille et un terme proportionnel à l'accroissement de taille a été proposé par Gascuel (1994) :

$$\sigma(t) = \alpha + \beta L(t) + \gamma \frac{dL(t)}{dt} \quad (4)$$

Où  $\alpha$ ,  $\beta$  et  $\gamma$  sont les paramètres du modèles,  $t$  est l'âge conventionnel et  $L(t)$  est la longueur moyenne à l'âge  $t$ .

Lorsque la loi de croissance utilisée est un modèle de Von Bertalanffy (1938), l'équation (4) peut être ré-exprimée en fonction de l'âge :

$$\sigma(t) = (\alpha + \beta L_{\infty}) + (\gamma K - \beta) \left( \exp(-K(t - t_0)) \right) \quad (5)$$

Si plusieurs années de données de fréquences de taille sont disponibles, un premier ajustement sans contrainte sur les écarts types permet d'estimer les paramètres du modèle (4). La prédiction du modèle est alors utilisée pour définir la contrainte sur les écarts types.

Dans la pratique il s'avère fréquent que le modèle (4) soit trop complexe et sur paramétré. En effet, le terme  $\beta \cdot L(t)$  correspond à l'idée que la variabilité des tailles augmente avec l'âge et donc la taille, tandis que le terme  $\gamma \cdot dL(t)/dt$  correspond à l'idée que les écarts de taille sont plus importants en période de forte croissance, c'est à dire pour les âges jeunes. Ces deux effets apparaissent donc contradictoires. L'expérience prouve que c'est généralement le second qui l'emporte et que contrairement à une idée très répandue les écarts de taille diminuent globalement avec la taille et l'âge des animaux. Dès lors, c'est un modèle simplifié négligeant le terme  $\beta \cdot L(t)$  qui doit être utilisé.

### 2.3. Données

Plusieurs types de données peuvent être utilisés lors de l'ajustement. En premier lieu, les données sur lesquelles l'ajustement du modèle de mélange est réalisé sont des données groupées, i.e. elles sont divisées en groupes plus ou moins larges auxquels sont associés les fréquences d'occurrence des individus. La taille des intervalles des groupes est importante car elle détermine la forme des histogrammes sur lesquels sont ajustés le modèle.

Les données conditionnelles sont des sous-échantillons fournissant de l'information complémentaire aux données groupées. Ce sont des échantillons pour lesquels on dispose de la composante (la classe) à laquelle appartient l'individu. Par exemple un échantillon d'otolithes de poissons permet de disposer de la taille et de l'âge de l'individu. L'ajustement qui résulte de l'incorporation de ces données conditionnelles aux données groupées est généralement meilleur.

Enfin, des données auxiliaires sont utiles pour définir les paramètres des distributions (moyennes et écarts-types), ainsi que leur nombre (par exemple le nombre de classes d'âge). Si l'inspection de l'histogramme des fréquences fournit une information utile pour les valeurs des paramètres initiaux des distributions, elle est souvent insuffisante pour déterminer le nombre de composantes et elle peut conduire à des résultats biaisés. On pourra ainsi utiliser une loi de croissance ou un modèle quelconque sur les paramètres, pour déterminer les valeurs de départ de l'estimation. Des données bibliographiques sur la population permettront souvent de déterminer le nombre de classes de la distribution mélangée.

### 2.4. Algorithmes numériques

La méthode statistique utilisée pour ajuster le modèle aux données est l'estimation par maximum de vraisemblance. Les algorithmes de calcul sont une méthode de type Newton et l'algorithme "Expectation-Maximization" (EM). Ce dernier algorithme est applicable à une grande variété de problèmes sur données incomplètes, pour lesquels il donne de meilleurs

résultats que les méthodes de type Newton. La fonction `nlm()` implémentée dans R est à la base de l'ajustement.

### 3. Cas d'étude : le bobo de Guinée

#### 3.1. Loi de croissance

Les paramètres de la loi de croissance de Von Bertalanffy (1938) ont été estimés pour le bobo par la méthode de Shepherd (1987) qui s'appuie sur des données de fréquences de taille moyennées sur la période 1995-2000, supposées représentatives d'une distribution des captures des pêches artisanale et industrielle de Guinée Conakry (Sidibé 2003; Fig. 1). Les paramètres  $L_{\infty}$  et  $K$  sont égaux respectivement à 48.63 cm et  $0.35 \text{ an}^{-1}$ . Dans cette analyse, la valeur du paramètre  $t_0$  a été modifiée par rapport à Sidibé (2003;  $t_0 = -0.051$ ) pour faire coïncider les pics observés dans les fréquences de taille avec les tailles moyennes de la courbe de croissance ( $t_0 = 0.05$ ).

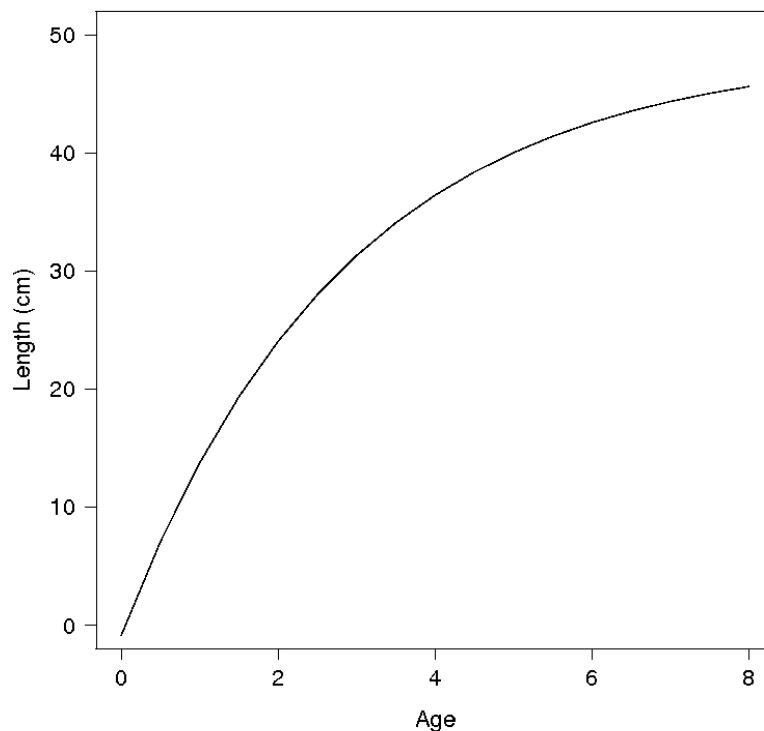


Fig. 1. Loi de croissance pour le bobo (*Pseudolithus elongatus*) estimée par la méthode de Shepherd (1987) pour le stock de Guinée (Sidibé 2003).

#### 3.2. Influence de la paramétrisation sur les ajustements

##### . Matériel et Méthodes

Dans un premier temps, 2 ajustements ont été réalisés en faisant une hypothèse de normalité des distributions de taille aux âges. Un premier ajustement a été fait sans contrainte sur les proportions, moyennes et écarts types. Un deuxième ajustement a été réalisé avec les tailles moyennes fixées aux valeurs calculées à partir de la loi de croissance (Fig. 1) sans contrainte sur les écarts types. Pour faciliter l'estimation des paramètres, les

valeurs d'initialisation des tailles moyennes ont été fixées aux valeurs dérivées de la courbe de croissance pour les âges 0.5 à 6.5.

### Résultats

Les lois de distribution estimées sont très différentes selon la paramétrisation utilisée (Fig. 2). Excepté pour l'âge initial où les 2 distributions sont proches, les distributions estimées pour les autres classes d'âge diffèrent de manière substantielle (Fig. 2). L'ajustement sans contrainte (#1) donne des distributions relativement cohérentes du point de vue biologique pour les âges 1-3 mais l'optimisation conduit à un résultat aberrant pour la classe d'âge 4 caractérisée par une valeur de fréquence très élevée et un écart type très faible (Fig. 2). Dans le deuxième ajustement (#2), fixer les tailles moyennes en laissant libres les écarts types conduit à des résultats incohérents sur le plan biologique avec par exemple les poissons de la classe d'âge 2 pouvant mesurer entre 0 et 35 cm alors que les individus d'âge 3 sont au contraire caractérisés par un écart type faible, i.e. des tailles comprises entre environ 23 et 33 cm (Fig. 2).

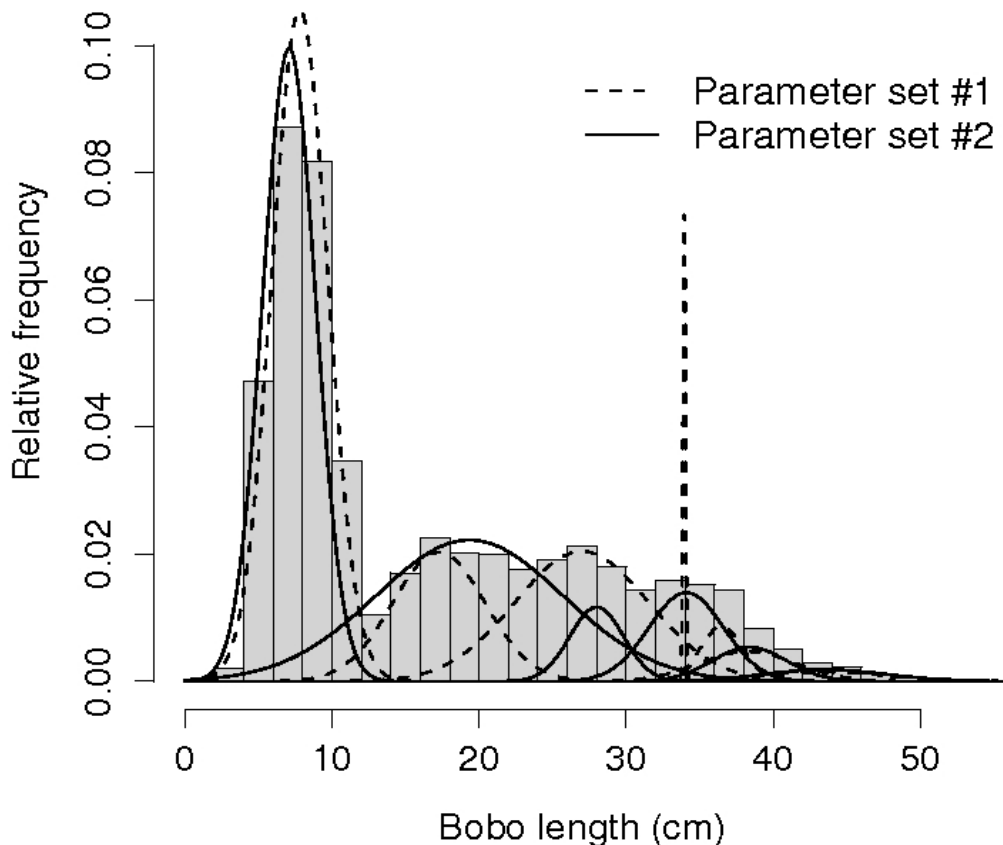


Fig. 2. Histogramme des fréquences de taille et lois de distribution des fréquences de tailles aux âges estimées sans contrainte (ligne pointillée) et sans contrainte (ligne continue; écarts type fixés).

L'initialisation des tailles moyennes et écarts types à des valeurs proches des résultats facilite le processus d'optimisation en partant d'une 'solution' statistique proche du maximum de vraisemblance. Les résultats suggèrent que l'estimation d'un 'optimum' statistique basé ici sur la théorie du maximum de vraisemblance ne conduisent pas toujours à des résultats cohérents du point de vue biologique. Dans ce contexte, l'utilisation d'une contrainte

supplémentaire sur les écarts types permet de rajouter de l'information biologique et d'aider à la décomposition des fréquences de tailles en fréquences d'âges.

### 3.3. Contrainte sur les écarts types

#### . Matériel et Méthodes

Les fréquences de tailles annuelles pour la pêche artisanale et industrielle de bobo pour la période 1995-2000 ont été utilisées pour définir une contrainte sur les écarts type des fréquences aux âges suivant la méthode de Gascuel (1994) en s'appuyant sur l'équation (5) définie pour une courbe de croissance de type Von Bertalanffy (1938). Des mélanges de lois normales ont été ajustés sur les jeux de données de fréquence de taille en excluant les fréquences de taille de la pêche industrielle pour l'année 1997 car la gamme de tailles a été considérée non représentative de la pêcherie, les tailles de bobo étant uniquement représentées de 5 à 17 cm (Sidibé 2003, p314). La taille moyenne aux âges a été fixée suivant la loi de croissance estimée par Sidibé (2003) et celles des écarts types aux âges ont été fixées aux valeurs estimées par Sidibé (2003, p 319). Un modèle linéaire a ensuite été ajusté pour les âges 2-6 car les écarts types estimés pour les âges 1 et 7 sont très variables du fait du manque d'observations pour les petits (< 10 cm) et grands poissons (> 45 cm).

#### . Résultats

Le modèle linéaire montre une décroissance linéaire significative de l'écart type des fréquences de tailles aux âges 2-6 ( $r$  ajusté = 0.72,  $p < 0.001$ ) (Fig. 3). Les résidus du modèle sont cohérents avec l'hypothèse de normalité malgré une valeur d'écart type éloignée du modèle pour l'âge 3 estimée à partir des données de fréquences de taille de la pêcherie artisanale en 1995.

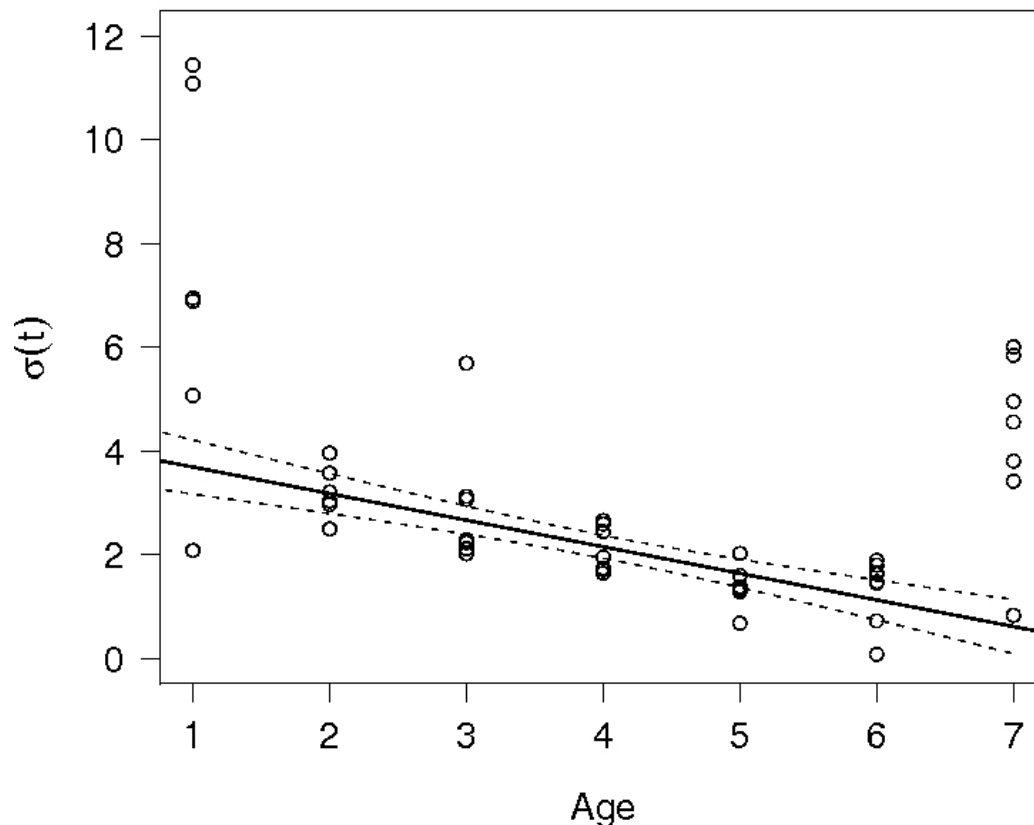


Fig. 3. Modèle de variabilité des écarts types des fréquences de tailles aux âges sous hypothèse de normalité en fonction de l'âge. Le modèle de régression linéaire a été ajusté pour les âges 2-6. Les lignes pointillées indiquent l'intervalle de confiance d'estimation à 95%.

### 3.4. Ajustement avec contrainte sur les écarts types

#### Matériel et Méthodes

Suivant le modèle linéaire reliant les écarts types des fréquences de taille aux âges (section 3.3), 2 nouveaux ajustements ont été réalisés en fixant les écarts types pour les âges 2-6 (paramétrisation #3) et 2-7 (paramétrisation #4), sans contrainte sur les proportions et les tailles moyennes.

#### Résultats

L'ajustement en fixant les écarts types pour les âges 2-6 est bon et conduit à des résultats cohérents sur le plan biologique sauf pour la classe d'âge 7 dont la moyenne est estimée à environ 40 cm, c'est à dire à la même valeur que pour la classe d'âge 6 (Fig. 4). L'écart type du dernier groupe d'âge est très important (5.05) et implique que les poissons de ce groupe ont des tailles comprises entre < 20 cm et > 50 cm. La présence de poissons pouvant être issus de plusieurs classes d'âges > 7 dans ce groupe + tend ainsi à l'estimation de grands écarts types lorsque aucune contrainte n'est considérée pour ce groupe. Dans ce cas, le choix d'une loi normale symétrique peut affecter les résultats pour les classes d'âge précédentes car la grande dispersion autour de la moyenne (due à quelques poissons de grande taille) implique qu'une certaine proportion des poissons ayant des tailles comprises entre 30 et 40 cm sont affectés à la classe d'âge 7, ceci affectant les lois estimées pour les classes d'âges 4-6 (Fig. 4).



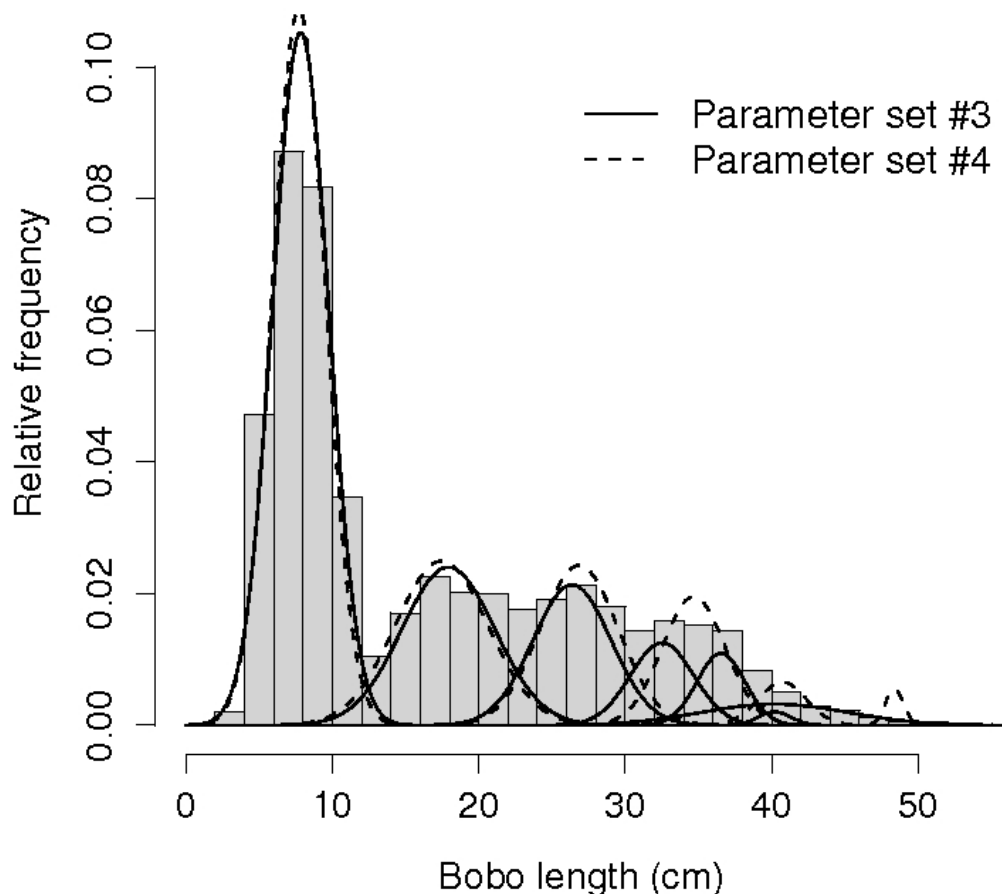


Fig. 4. Histogramme des fréquences de taille et lois de distribution des fréquences de tailles aux âges estimées avec contrainte sur les âges 2-6 (ligne continue) et sur les âges 2-7 (ligne pointillée).

Lorsqu'une contrainte est ajoutée sur l'écart type du dernier groupe d'âge (paramétrisation #4), l'ajustement devient moins bon pour les plus grandes tailles (résidus plus importants que pour la paramétrisation #3), mais les résultats apparaissent plus cohérents du point de vue biologique. Dans ce cas, les tailles moyennes estimées sont très proches de celles calculées à partir de la loi de croissance estimée par Sidibé (2003) excepté pour la dernière classe d'âge avec une taille moyenne estimée d'environ 48.4 cm par rapport à 43.5 cm pour le modèle de Sidibé (2003). Cette grande taille moyenne pour le dernier groupe d'âge suggère le fait que ce groupe peut inclure des poissons d'âge 7 et plus, i.e. un groupe d'âge +. Dans cette situation, fixer l'écart type pour la classe 'âge 7+' en extrapolant le modèle estimé pour les âges 2-6 semble donner des résultats moins satisfaisants du point de vue statistique (moins de degrés de liberté pour le modèle) mais plus cohérents en termes biologiques.

## 4. Discussion

### 4.1. R et mixdist

La disponibilité de données de lectures directes d'âge ou de données de marquage est généralement rare dans les pays d'Afrique de l'ouest du fait : 1) des difficultés d'âgeage de nombreuses espèces tropicales, 2) du coût et des moyens importants associés à l'obtention de ce type de données dans les pays en voie de développement. Dans ce contexte, le recours à des techniques de décomposition polymodale pour passer de données de fréquences de taille en données de fréquences d'âge constitue une phase initiale clé pour

toute application de modèle structuré en âge. La surparamétrisation caractéristique de ce type de modèles de mélanges implique généralement de fixer des contraintes sur les paramètres à estimer pour combiner critères statistiques et réalité biologique lors de l'estimation.

La librairie 'mixdist' développé sous R (Du 2002) offre un cadre flexible pour modéliser les données de fréquences de tailles et estimer les distributions sous-jacentes à partir d'un algorithme de calcul puissant et adapté. L'intérêt de la plateforme de modélisation R (R Development Core Team 2008) est dans ce cas double. D'une part, R est libre de droits et peut être installé localement sur un ordinateur sans aucun frais pour son utilisateur. D'autre part, R permet de développer des chaînes d'analyse, allant de l'extraction des données, à leur visualisation, leur traitement par des méthodes statistiques et de modélisation et enfin à la visualisation des résultats et leur intégration dans des rapports et articles scientifiques. Ainsi, la philosophie retenue au cours du projet ISTAM est de développer plusieurs modules, de l'amont avec l'extrapolation des données par exemple, à l'aval avec l'utilisation de modèles d'évaluation tels que l'analyse rectifiée conduite sur pseudo-cohorte et de pouvoir in fine intégrer toutes ces analyses via R. La disponibilité des scripts R dans les annexes des différents livrables produits doit offrir aux scientifiques présents dans les groupes de travail du COPACE et les Commissions une base pour appliquer ces méthodes à d'autres stocks.

#### **4.2. Exemple du bobo de Guinée**

L'exemple du bobo de Guinée a permis de mettre en évidence l'intérêt de la méthode de décomposition polymodale pour une espèce de poisson démersal pour laquelle plusieurs jeux de données de fréquences de taille avaient été collectés pour différents segments de la pêcherie de bobo en Guinée Conakry. La méthode dont les scripts R sont donnés en annexe est simple d'utilisation et peut être transposée à n'importe quelle espèce pour laquelle des fréquences de tailles sont disponibles. Pour ce type d'approche, il est néanmoins nécessaire qu'un minimum d'information soit contenue dans les données, i.e. en particulier que des pics correspondant aux cohortes (période de recrutement bien marquée) rentrant dans la pêcherie apparaissent dans les fréquences de taille.

Un intérêt majeur de la librairie 'mixdist' est de proposer un ensemble de contraintes à ajouter lors de l'ajustement qui sont faciles à implémenter car définies comme des paramètres de la fonction 'mix'. L'exemple du bobo a en outre permis de présenter la méthode de Gascuel (1994) et de développer le code R pour cette méthode pour ajouter une contrainte sur les écarts types des fréquences de tailles aux âges. La fixation des écarts types à partir de leur modélisation en fonction de l'âge a permis l'estimation des moyennes de ces fréquences de tailles (qui peuvent différer des tailles moyennes estimées à partir de la courbe de croissance du stock) tout en obtenant des résultats cohérents du point de vue biologique, les gammes de taille comprises dans chaque classe d'âge devant être réalistes.

## **Références**

- Du, J. 2002. Combined algorithms for constrained estimation of finite mixture distributions with grouped data and conditional data. McMaster University MS thesis, 137p.
- Fridriksson, A. 1934. On the calculation of age distribution within a stock of cod by means of relatively few age-determination as a key to measurements on a large scale. Rapp. P.-v. Réun. Cons. perm. Inter. Explor. Mer 86 : 1-14.

- Gascuel, D. 1994. Une méthode simple d'ajustement des clés taille/âge : application aux captures d'albacore (*Thunnus albacares*) de l'Atlantique Est. Canadian Journal of Fisheries and Aquatic Sciences 51 : 723-733.
- R Development Core Team (2008). R: A language and environment for statistical computing, reference index version 2.2.4. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rouyer, T. 2006. Le package mixdist sous R. Pour la décomposition polymodale. Rapport interne non publié, Agrocampus, 12p.
- Shepherd, J.G. 1987. A weakly parametric method for estimating growth parameters from length composition data. In : Pauly, D., Morgan, G.R. (eds), Length-based methods in fisheries research. ICLARM Conf. Proc. 13 : 113-119.
- Sidibé, A. 2003. Les ressources halieutiques démersales côtières de la Guinée : exploitation, biologie et dynamique des principales espèces de la communauté à Sciaenidés. Thèse de Doctorat Halieutique, Ensa-Rennes, 320 p.
- Von Bertalanffy, L. 1938. A quantitative theory of organic growth (Inquiries on growth Laws II). Human Biology 10 : 181-213.

## Annexe – Scripts R

```
rm(list = ls())
rep.data <- Folder with data
rep.res <- Folder where results are stored
setwd(rep.data)

#####
### Growth curve ###
#####

### Length-weight parameters p140 (Sidibé 2003)
a <- 0.0058
b <- 3.123

### Growth parameters

Linf <- 48.63 #p143 (Sidibé 2003)
K <- 0.35
t0 <- 0.05 #original value -0.051
n<-7 # number of age-classes

### Computation of mean lengths Lt
i<-seq(0.5,n-0.5,by=1)
L<-Linf*(1-exp(-K*(i-t0)))

#####
### Plotting the growth curve
#####
age.sim <- seq(0,8,by = 0.5) #simulates ages from 0 to 8 years
size <- Linf*(1-exp(-K*(age.sim-t0))) #computes the growth from the growth model

setwd(rep.res)
postscript('bobo-growth.ps',pointsize=12,height=8,width=8, horizontal = FALSE)
plot(size ~ age.sim, xlim = c(0,8), ylim=c(0,50), xlab="Age", ylab = "Length (cm)", las= 1, type
= "l", cex.lab = 1.3, cex.axis = 1.3)
dev.off()
setwd(rep.data)

#####
### Data
#####
library(mixdist)
data_brut<-read.table('data-bobo1.txt',header=T)
attach(data_brut)
data <- rep(data_brut$taille,data_brut$freq) #produces the 'real' vector of data observed
databobo<-mixgroup(data,breaks=c(0,seq(2,52,2),54),xname="Taille")

### Fitting with 2 different sets of parameters
### Initialisation
sigma1 <- c(2.95,2.8,1.99,1.55,1.4,1.25,3.1) #values from Sidibé (2003)
parambobo1 <- mixparam(L,sigma1) #parameters for the mixed model

# Alternative parametrization
sigma2 <- c(2,2,2,2,2,2,3) #other values (arbitrary) for sigma
```

```

parambobo2 <- mixparam(L,sigma2)

mixbobo1<-
mix(databobo,parambobo1,dist="norm",emsteps=100,mixconstr(conmu="MFX",fixmu=c(T,T,
T,T,T,T),consigma="SFX",fixsigma=c(F,F,F,F,F,F)))      #fitting function

mixbobo2<-
mix(databobo,parambobo2,dist="norm",emsteps=100,mixconstr(conmu="MFX",fixmu=c(F,F,
F,F,F,F,F),consigma="SFX",fixsigma=c(F,F,F,F,F,F)))      #fitting function

#Get the results
mean1 <- mixbobo1$parameters$mu
prop1 <- mixbobo1$parameters$pi
sd1 <- mixbobo1$parameters$sigma

mean2 <- mixbobo2$parameters$mu
prop2 <- mixbobo2$parameters$pi
sd2 <- mixbobo2$parameters$sigma

#Figure
#No constaint vs. fixed means
setwd(rep.res)
postscript('mix-bobo-free-comp.ps',pointsize=15,horizontal=F,height=7,width=9)
opar <-par()
par(mar = c(5,5,1,2))
hist(data,breaks = seq(0,54,2),xlab = "Bobo length (cm)", ylab = "Relative frequency", main =
"",col = "lightgrey", freq = FALSE, cex.axis = 1.3, cex.lab = 1.3, ylim = c(0,0.11))

lgen <- seq(0,60, by = 0.01)
for (i in (1:n)){
dens1 <- dnorm(lgen,mean = mean1[i],sd = sd1[i])
dens.rel1 <- dens1/sum(dens1) * prop1[i] * 100
lines(lgen,dens.rel1,col = "black", type = "l", lwd = 2)

dens2 <- dnorm(lgen,mean = mean2[i],sd = sd2[i])
dens.rel2 <- dens2/sum(dens2) * prop2[i] * 100
lines(lgen,dens.rel2,col = "black", type = "l", lwd = 2, lty = 2)

}
legend(x = 25, y = 0.1, legend = c("Parameter set #1", "Parameter set #2"), lty = c(2,1), lwd =
c(2,2), bty = "n", cex = 1.3)
dev.off()
par(opar)

### Parameter estimation of the relationship  $\sigma(t)=\alpha+\beta*L(t)+\gamma*dL(t)/dt$ 
# through a fitting without constraint on size frequency histograms for all years available
setwd(rep.data)
freqtaille <- read.table('freqtaille-bobo2.csv',header=T, sep = ",")
res_sd <- matrix(NA, nrow = n, ncol = ncol(freqtaille)-1)

for (i in (2:ncol(freqtaille)))
{
size <- rep(na.omit(freqtaille[,1]),na.omit(freqtaille[,i]))
datamixbobo <- mixgroup(size,breaks=c(0,seq(2,52,2),54),xname="Taille")

```

```

        mixbobo <- mix(datamixbobo,parambobo1,dist = "norm",emsteps =
100,mixconstr(conmu = "MFX",fixmu = c(T,T,T,T,T,T,T),consigma = "SFX",fixsigma =
c(F,F,F,F,F,F,F)))    #fitting function
        res_sd[i,i-1] <- mixbobo$parameters$sigma
    }

#Fitting a linear model
#For ages 2-6
age <- 2:6
data_sd <- data.frame(res_sd[age,-7])    #selection of ages 2-6 (removing of industrial
fishing samples from 1997) and conversion of the matrix into a df
seq_age <- rep(2:6,8)    #ages of the df
data_lm_sd <- cbind(seq_age,stack(data_sd))[-3] #values of sd of the df
std_bobo <- data_lm_sd[,2]
age_bobo <- data_lm_sd[,1]
lmsd <- lm(std_bobo~age_bobo)    #linear regression
anova(lmsd)
sum_sd <- summary(lmsd)
sum_sd$coefficients

###Figure constraint for the sd
setwd(rep.res)
postscript('sigma-compute.ps',pointsize=15,horizontal=F,height=7,width=9)
plot(1:n,res_sd[(1:n),1],ylim = c(0,12), las = 1, cex.axis = 1.3, cex.lab = 1.3, xlab = "Age",
ylab = expression(sigma(t)))
for (i in c(2:6,8:9)){
points((1:(n)),res_sd[(1:(n)),i])
}
#points(1:n,res_sd[1:n,7], col = "red")    #Industrial fishing in 1997

#Adding the linear model to the plot
newd <- data.frame(age_bobo = seq(0,7, by = 0.05))    #generates age data
pred.sd <- predict(lmsd,newd,interval = "confidence")    #predicts the value of sd based on
the linear regression model lmsd
abline(a = sum_sd$coefficients[1], b = sum_sd$coefficients[2],lwd = 2)
points(pred.sd[,2]~newd$age_bobo, type = "l", lty = 2)
points(pred.sd[,3]~newd$age_bobo, type = "l", lty = 2)
dev.off()

#Plot the residuals
setwd(rep.res)
postscript('residuals.ps',pointsize=15,horizontal=F,height=7,width=9)
plot(age_bobo,residuals(lmsd), ylim = c(-10,10), xlab = "Age", ylab = "Residuals", cex.lab =
1.3, cex.axis = 1.3)
abline(h = 0, lwd = 2, lty = 2)
dev.off()

#####
## Fitting with constraint on the SDs derived from the linear model
#####
setwd(rep.data)

```

```

#Get the coefficients
age_pred <- data.frame(age_bobo = seq(1,7, by = 1)) #ages 1-7
sigma3 <- predict(lmsd,age_pred)$fit #estimates sd at ages from the linear model
parambobo3 <- mixparam(L,sigma3) #new parameters for the mixture model

#Fixing SDs for ages 2-6
mixbobo3 <-
mix(databobo,parambobo3,dist="norm",emsteps=100,mixconstr(conmu="MFX",fixmu=c(F,F,
F,F,F,F),consigma="SFX",fixsigma=c(F,T,T,T,T,T,F))) #fitting function

#Fixing SDs for ages 2-7
mixbobo4 <-
mix(databobo,parambobo3,dist="norm",emsteps=100,mixconstr(conmu="MFX",fixmu=c(F,F,
F,F,F,F),consigma="SFX",fixsigma=c(F,T,T,T,T,T,T)))

#Get the results
mean3 <- mixbobo3$parameters$mu
prop3 <- mixbobo3$parameters$pi
sd3 <- mixbobo3$parameters$sigma

mean4 <- mixbobo4$parameters$mu
prop4 <- mixbobo4$parameters$pi
sd4 <- mixbobo4$parameters$sigma

#Figure
setwd(rep.res)
postscript('mix-bobo-constr-comp.ps',pointsize=15,horizontal=F,height=7,width=9)
opar <-par()
par(mar = c(5,5,1,2))
hist(data,breaks = seq(0,54,2),xlab = "Bobo length (cm)", ylab = "Relative frequency", main =
"",col = "lightgrey", freq = FALSE, cex.axis = 1.3, cex.lab = 1.3, ylim = c(0,0.11))

lgen <- seq(0,60, by = 0.01)
for (i in (1:n)){
dens3 <- dnorm(lgen,mean = mean3[i],sd = sd3[i])
dens.rel3 <- dens3/sum(dens3) * prop3[i] * 100
lines(lgen,dens.rel3,col = "black", type = "l", lwd = 2)

dens4 <- dnorm(lgen,mean = mean4[i],sd = sd4[i])
dens.rel4 <- dens4/sum(dens4) * prop4[i] * 100
lines(lgen,dens.rel4,col = "black", type = "l", lwd = 2, lty = 2)

}
legend(x = 25, y = 0.1, legend = c("Parameter set #3", "Parameter set #4"), lty = c(1,2), lwd =
c(2,2), bty = "n", cex = 1.3)
dev.off()
par(opar)

```