

**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 – Document **xx**

**Analysis of CPUE data series through Generalized Linear Models and Delta method to derive annual series of abundance indices**

**Application to the Mauritanian demersal fishery**

**November 2008**

**Beyah O. MEISSA <sup>(1)</sup>, Etienne RIVOT <sup>(2,3)</sup> and Didier GASCUEL <sup>(2,3)</sup>**

<sup>(1)</sup> Institut Mauritanien de Recherches Océanographiques et des Pêches (IMROP), Nouadhibou, Mauritanie.

<sup>(2)</sup> Aquatic and Fisheries Sciences Center, Agrocampus OUEST, Rennes, France

<sup>(3)</sup> Laboratory of Biostatistic and Fishery Information System. INRH, National Institute of fisheries sciences, Casablanca, Morocco

## ABSTRACT

This scientific technical report stands in the Working Package 3 “*Stock Assessment Methods and Analysis Tools*” of the ISTAM European project (<http://projet-istam.org>). The work package WP3 aims at improving the stock assessment methods, with a particular emphasis to models developed in data poor environment and for short lived species.

The present work presents an application of Generalized Linear Models to derive series of abundance indices for different species of the demersal fishery in Mauritania.

Estimate of abundance indices is a key step of stocks assessment. In west Africa such indices are frequently estimate based on methods which are not statistically optimum. As a consequence, they can be biased and/or exhibit a high uncertainty. The present report deals with an application of the Generalized Linear Models, as a robust statistical method used to derive series of abundance indices from data surveys. The method is summarized and explained for non statistician users and the method is applied as a test to a set of various species of the demersal fishery in Mauritania. The related R package is presented in an appendix of this report.

## CONTENTS

1.	Introduction.....	2
2.	Méthode <i>delta-GLM</i> .....	2
2.1.1.	Calcul des CPUE .....	2
2.1.2.	Principes généraux de la méthode .....	3
2.1.3.	Sous modèle 1 : Distribution binomiale de présence-absence .....	3
2.1.4.	Sous modèle 2 : Distribution gaussienne du log des CPUE pour les observations non nulles .....	4
2.1.5.	Combinaison des deux sous-modèles .....	4
3.	Résultats.....	5
3.1.	Thiof ( <i>Epinephelus aeneus</i> ).....	5
3.1.1.	Distribution des données de CPUE .....	5
3.1.2.	Modèle avec distribution binomiale pour la présence-absence.....	5
3.1.3.	Modèle log-gaussien pour les CPUE >0 .....	6
3.1.4.	Les indices d'abondances estimés .....	6
3.2.	Courbine ( <i>Argyrosomus regius</i> ).....	7
3.3.	Diagramme ( <i>Plectorinchus mediterraneus</i> ) .....	8
3.4.	Saint Pierre ( <i>Zeus faber</i> ) .....	9
3.5.	Ombrine ( <i>Umbrina canariensis</i> ).....	9
4.	Discussion & conclusions .....	9
4.1.	Limites et sources de biais des CPUE comme indices d'abondances.....	9
4.2.	Limites des indices issus des campagnes scientifiques .....	9
4.3.	Conclusions .....	9
5.	ANNEXE : Script R du modèle GLM (cas du thiof).....	11

## 1. INTRODUCTION

Le calcul d'indices d'abondance est une étape clés de l'évaluation des stocks. De tels indices présentent un intérêt intrinsèque car ils renseignent sur les tendances d'évolution de la ressource. Ils sont également à la base de la plupart des modèles de dynamique des populations, notamment pour le calcul d'efforts de pêche théoriques et pour l'ajustement des modèles de production, ou pour la calibration des méthodes d'analyse des cohortes. Ces indices sont généralement estimées à partir des données d'observations issues des campagnes scientifiques, ou à partir des statistiques de pêches issues des flottilles commerciales. Dans un cas comme dans l'autre, estimer des indices d'abondance reste cependant une tâche relativement complexe, compte tenu notamment de plan d'échantillonnage souvent fortement déséquilibrés. Dès lors, force est de constater que bien souvent les méthodes utilisées ne sont pas statistiquement optimales, ce qui peut conduire à des estimations biaisées et/ou à des incertitudes très fortes.

Dans le présent travail une méthode statistique rigoureuse basée sur les Modèles Linéaires Généralisés (GLM) est présentée. A titre d'illustration, elle est mise en œuvre pour estimer une série annuelle d'indices d'abondance standardisés, à partir des captures par unités d'effort (CPUE) de la flotte nationale Mauritanienne, ou à partir des indices d'abondances issus des campagnes scientifiques. Les indices d'abondances sont ici estimés pour une sélection de huit espèces et la cohérence des deux séries d'indices (campagnes scientifique et données commerciales) est brièvement analysée.

De nombreuses co-variables sont susceptibles de faire varier la capture par unité d'effort : type du bateau, puissance du bateau, zone de pêche ... . La procédure statistique mise en œuvre permet de faire le choix des variables explicatives à conserver dans le modèle statistique. Le GLM "rééquilibre" en quelque sorte, le plan d'échantillonnage pour toutes les modalités des variables qui sont retenues dans le modèle. Les variables retenues sont celles qui ont un effet statistiquement significatif et qui expliquent une part non négligeable de la variabilité des données. Le principe consiste alors à filtrer l'effet des variables retenues autre que l'année, pour extraire une série d'indice d'abondance annuel.

## 2. METHODE *DELTA-GLM*

### 2.1.1. Calcul des CPUE

Pour les CPUE des captures commerciales, les indices d'abondances sont calculés à partir des captures par unité d'effort (CPUE) des flottilles commerciales nationales débarquant en Mauritanie. Les captures par unité d'effort (CPUE) sont obtenues par la division des débarquements par espèce et par marée sur l'effort de pêche déployé (en heures de pêche) pour réaliser cette capture, soit :

$$(1) \quad CPUE = \frac{C_{s,n}}{E_{s,n}}$$

où  $C_{s,n}$  est la capture réalisée de l'espèce  $s$  lors de la marées  $n$  et  $E_{s,n}$  l'effort de pêche total correspondant.

Pour les CPUE issues de campagnes scientifiques, l'effort est généralement standardisé dès la phase d'observation (chalut et durée du trait standardisés). Le calcul de la CPUE correspondant à chaque trait de chalut reste analogue à l'équation (1). L'estimation d'indices d'abondance revient ici à calculer des rendements standardisés par demi-heure de chalutage. Le GLM utilisé prend donc en compte les facteurs susceptibles d'affecter les rendements de pêche scientifiques, notamment la saison et la zone de pêche (en latitude et en bathymétrie).

Deux séries d'indices d'abondance basées sur les données scientifiques ont été récemment publiées, l'une exprimée en rendement par demi-heure de chalutage (Gascuel et al., sous presse), l'autre en biomasse totale (Gascuel et al., 2007). Cette deuxième série tient compte des différences de densité entre strates bathymétriques ou latitudinales et de la surface de chaque strate ; elle est ainsi considérée plus fiable que la première et sera la seule utilisée ici. Notons que ces indices sont calculés en sélectionnant les campagnes du plateau continental dont le nombre de traits est supérieur à 20. Les données sont ensuite agrégées par campagne, par zone (Nord, Centre et Sud) et par strate bathymétrique pour réduire le nombre d'observations nulles liées principalement à la faible surface chalutée. L'analyse des données de campagnes scientifiques prend aussi en compte une standardisation des données par rapport à la modification apportées sur l'engin de pêche en 1989. Des coefficients de corrections ont aussi été calculés (Gascuel et al., 2007), à partir des données de campagnes d'inter-calibration rapportées par Girardin et al. (1990).

### 2.1.2. Principes généraux de la méthode

Les GLM sont une extension du Modèle linéaire (LM) dont l'utilisation est limitée à certaines conditions, notamment la normalité des résidus, condition rarement respectée dans l'analyse des données halieutiques. Cependant, le principe reste le même que celui de l'analyse par modèle LM et consiste à faire le lien entre des variables à expliquer (réponses, ici la CPUE) et des variables explicatives (l'année, la saison, la catégorie de navire, ...).

Plus précisément, la modélisation des CPUE commerciales permet de prendre en compte, d'une part, les variations spatio-temporelles de la disponibilité de la ressource, et d'autre part, les différences de puissance de pêche liées aux caractéristiques des navires (potentiellement : la puissance motrice du navire, son type de conservation et sa licence de pêche). On teste donc les effets des facteurs suivant : effet *An*, effet *saison*, effet *mois*, effet *TJB*, effet *type de conservation* (glacier ou congélateur) et effet *licence de pêche*. Certains facteurs sont testés avec différentes stratifications.

L'importance généralement élevée des enregistrements nuls dans les CPUE empêche souvent l'utilisation d'un modèle de type LM. Pour remédier à cela, on a recour ici à un modèle de type *Delta*. (in : Le Pape et al. 2003 ; Laurans 2005). Le modèle est obtenu en combinant deux sous-modèles : *i*) un premier modèle pour modéliser l'absence et la présence de l'espèce dans les traits de chalut ; *ii*) l'autre pour modéliser les valeurs positives des CPUE.

La méthodologie décrite ci-dessous est reconduite pour 8 espèces différentes, et pour l'analyse des données de CPUE commerciales et de campagnes scientifiques : *Epinephelus aeneus* (thiof), *Argyrosomus regius* (courbine), *Plectorhinchus mediterraneus* (diagramme), *Zeus faber* (Saint-Pierre), *Umbrina canariensis* (ombrine), *Epinephelus alexandrinus* (badeche), *Pomadasys incisus* (grondeur) et *Uranoscopus sp* (rata).

### 2.1.3. Sous modèle 1 : Distribution binomiale de présence-absence

Les probabilités de présence sont obtenues en premier temps par un recodage binaire (0 = absence et 1 = présence) des CPUE. Ensuite, un modèle GLM basé sur l'hypothèse d'une distribution binomiale est mis en œuvre :

$$(2a) \quad Z_{i,m,n} = \text{logit}(p_{i,m,n}^{>0})$$

$$(2b) \quad Z_{i,m,n} = \alpha_j + \beta_k + \gamma_l + \delta_m$$

où  $P_{i,j,k,l,m}$  est l'espérance de la probabilité de présence de l'espèce considérée pour le bateau *i* (qui appartient au type de navire *k* et à la classe de taille *j*), calculée pour l'année *m* et de la marée *n* (qui appartient à la saison *l*). Avec :  $\alpha_j$  un effet classe de taille de bateau exprimée en *TJB*,  $\beta_k$  un effet de type de navire (congélateur ou glacier),  $\gamma_l$  un effet de saisonnalité,  $\delta_m$  un effet année.

Respectivement  $j$ ,  $k$ ,  $l$  et  $m$  sont utilisés pour indiquer les niveaux de chaque variable. Notons que ce modèle est écrit ici avec seulement 4 variables explicatives et que dans les analyses il peut être élargi à d'autres variables explicatives existantes dans les bases de données.

La prédiction de la probabilité de présence (pour la catégorie de navire  $k$ , de classe de taille  $l$ , au cours de la saison  $l$  de l'année  $m$ ) est faite en premier lieu dans l'échelle de la fonction de lien  $\text{logit}()$  et le passage au référentiel initial est obtenu par la transformation inverse (calcul automatique sous R) :

$$(3) \quad \hat{p}_{j,k,l,m}^{>0} = \frac{e^{\hat{z}_{j,k,l,m}}}{1 + e^{\hat{z}_{j,k,l,m}}}$$

#### 2.1.4. Sous modèle 2 : Distribution gaussienne du log des CPUE pour les observations non nulles

Ici un modèle GLM est appliqué uniquement sur les enregistrements correspondant à une CPUE positive.

Dans un premier temps, une transformation logarithmique est appliquée sur les données afin d'homogénéiser les variances et de tenir compte de la nature multiplicative des effets étudiés. On suppose que la réponse  $\log(IA)$  suit un modèle linéaire du type :

$$(4) \quad U_{i,m,n} = \log(IA_{i,m,n}) = \alpha_j + \beta_k + \gamma_l + \delta_m + \varepsilon_{i,m,n}$$

où  $IA_{i,m,n}$  est la capture par unité d'effort du bateau  $i$  (qui appartient au type de navire  $k$  et à la classe de taille  $j$ ), calculée pour l'année  $m$  et de la marée  $n$  (qui appartient à la saison  $l$ ).  $\alpha_j$  est un effet classe de taille de bateau exprimé en TJB,  $\beta_k$  est un effet de type de navire (congélateur ou glacier),  $\gamma_l$  est un effet de saisonnalité,  $\delta_m$  est un effet année, et  $\varepsilon_{i,m,n}$  un terme d'erreur résiduelle. Les indices  $j$ ,  $k$ ,  $l$  et  $m$  sont utilisés pour indiquer les niveaux de chaque variable. Notons que ce modèle est écrit ici avec seulement 4 variables explicatives et que dans les analyses il peut être élargi à d'autres variables explicatives existantes dans les bases de données.

Afin de revenir à l'échelle initiale des CPUE, une transformation logarithmique inverse est effectuée en tenant compte de la correction de Laurent qui permet d'obtenir des valeurs non biaisées de l'espérance de l'indice d'abondance ( $IA$ ) :

$$(5) \quad \hat{IA}_{j,k,l,m}^{>0} = e^{\hat{U}_{j,k,l,m}^{>0}} \cdot e^{1/2 \cdot \sigma^2}$$

#### 2.1.5. Combinaison des deux sous-modèles

Les indices d'abondances prédits par le modèle delta ( $IA$ ) sont obtenus par le produit des probabilités de présences et des CPUE prédites pour les observations non nulles :

$$(6) \quad \hat{IA}_{j,k,l,m} = \hat{p}_{j,k,l,m}^{>0} \cdot \hat{IA}_{j,k,l,m}^{>0}$$

Les valeurs prédites par l'équation (6) sont calculées dans un tableau généré par une boucle sous R, pour toutes les combinaisons possibles des modalités des variables retenues. Ensuite, une moyenne sur l'ensemble des modalités des facteurs autres que l'année permet d'obtenir la prédiction de l'indice d'abondance pour l'année correspondante.

### 3. RESULTATS

Le calcul des indices d'abondance à partir des CPUE s'est porté sur les espèces suivantes : *Epinephelus aeneus* (thiof), *Argyrosomus regius* (courbine), *Plectorhinchus mediterraneus* (diagramme), *Zeus faber* (Saint-Pierre), *Umbrina canariensis* (ombrine), *Epinephelus alexandrinus* (badeche), *Pomadasys incisus* (grondeur) et *Uranoscopus sp* (rata). Seuls les résultats des cinq premières espèces seront présentés ici. Le choix des espèces modélisées est indirectement dicté par la base de données des débarquements dans laquelle ces espèces sont bien identifiées (elles ne sont pas regroupées avec d'autres espèces) et leur capture spécifique est connue dans cette base.

Les indices d'abondances des campagnes scientifiques, calculés par ailleurs, sont présentés en parallèle avec les indices issus des CPUE pour les espèces dont les deux indices sont calculés.

#### 3.1. Thiof (*Epinephelus aeneus*)

##### 3.1.1. Distribution des données de CPUE

L'histogramme des CPUE du thiof (figure 1) montre une proportion de valeurs nulles (0). Cela justifie l'utilisation d'une loi « Delta » qui caractérise le fait que pour un grand nombre de marées aucune capture n'est réalisée.

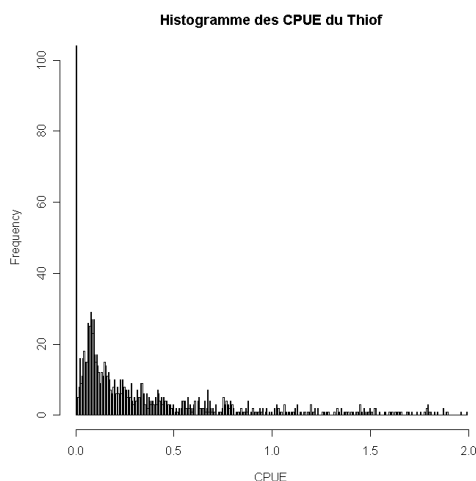


Figure 1 : Fréquences des valeurs de CPUE du thiof

##### 3.1.2. Modèle avec distribution binomiale pour la présence-absence

Pour le sous modèle présence/absence les facteurs retenus sont : l'année, la classe de TJB, le mois et le type de conservation. Le modèle explique 25% environ de la déviance. La variable année seule explique plus de 50% de la variabilité expliquée par le modèle.

Les probabilités de présence estimées pour le thiof montrent qu'il est de moins en moins présent dans les débarquements au cours de la série chronologique. Sa probabilité de présence au début des années 90 fluctuait entre 0,3 et 0,6, et elle est aujourd'hui de seulement 0,1 environ (soit 1 marée sur 10 par an).

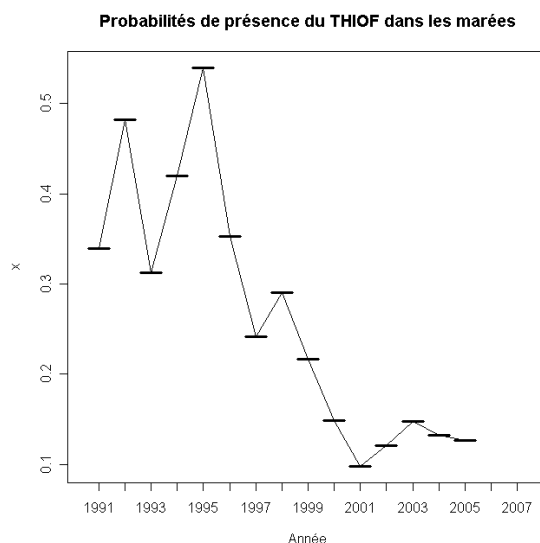


Figure 2 : Evolution des probabilités de présence de thiof dans les marées

### 3.1.3. Modèle log-gaussien pour les CPUE >0

Le modèle retenu intègre les facteurs explicatifs suivants : l'année de la marée, la classe de TJB du navire et son type de conservation (congélateur ou glacier). Le modèle explique 24% environ de la déviance. Le TJB et l'année expliquent la part la plus importante de la variabilité.

Les résidus en fonction des valeurs observées du modèle retenu sur les valeurs positives, montre globalement un bon ajustement avec l'absence d'autocorrélation des résidus, et une distribution assez homogène sur l'ensemble des valeurs. La droite d'Henry (Q-Q plot) qui compare les valeurs des quantiles de la loi empirique aux quantiles de la loi normale centrée réduite, confirme aussi la normalité de la distribution.

### 3.1.4. Les indices d'abondances estimés

L'indice d'abondance des campagnes scientifiques montre une synchronisation assez nette avec la CPUE commerciale, mise à part le pic observé en 2003 dans les campagnes scientifiques (figure 3). Selon les deux séries d'abondance, le stock de thiof serait en déclin sur toute la période exception faite pour les premières années. La diminution la plus forte de l'abondance du Thiof a été entre 1995 et 1998, ce qui coïncide avec la plus importante augmentation de l'effort de pêche (l'entrée en vigueur des accords de pêche entre la Mauritanie est l'UE).

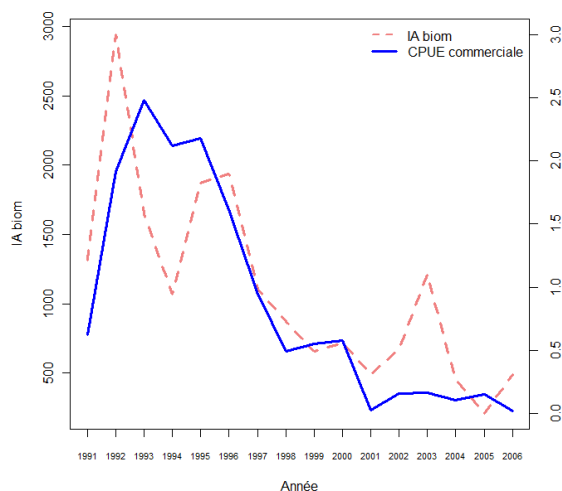


Figure 3 : Indices d'abondance standardisés du thiof : IA exprimés en rendement et en biomasse à partir des campagnes scientifiques (à gauche) ; avec la CPUE (à droite)

### 3.2. Courbine (*Argyrosomus regius*)

L'effet de saisonnalité est très marqué chez la courbine ; selon l'analyse des coefficients estimés et des probabilités de présence dans les marées par mois, il s'avère que les captures les plus importantes de cette espèce se font à partir de novembre jusqu'au mois d'avril (figure 4). La présence de cette espèce dans les captures des navires nationaux est faible en août sur toute la période étudiée.

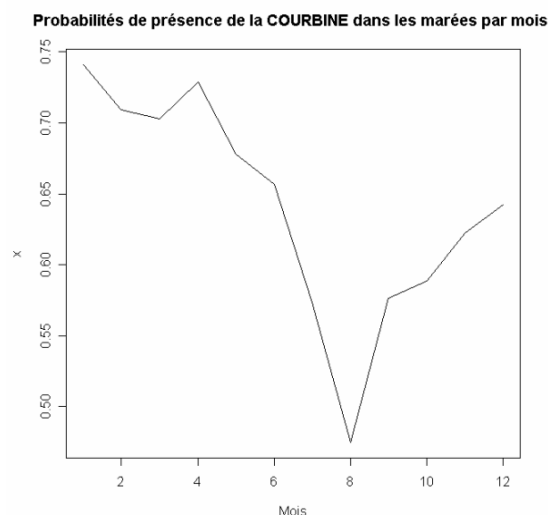


Figure 4 : Evolution saisonnière des probabilités de présence de l'*Argyrosomus regius* (courbine) dans les marées

Tandis que l'indice d'abondance issu des CPUE commerciales montre une baisse depuis 1997, celui des campagnes scientifiques fluctue sur toute la période et ne semble baisser qu'à partir de 2003 après le pic le plus élevé (figure 5).



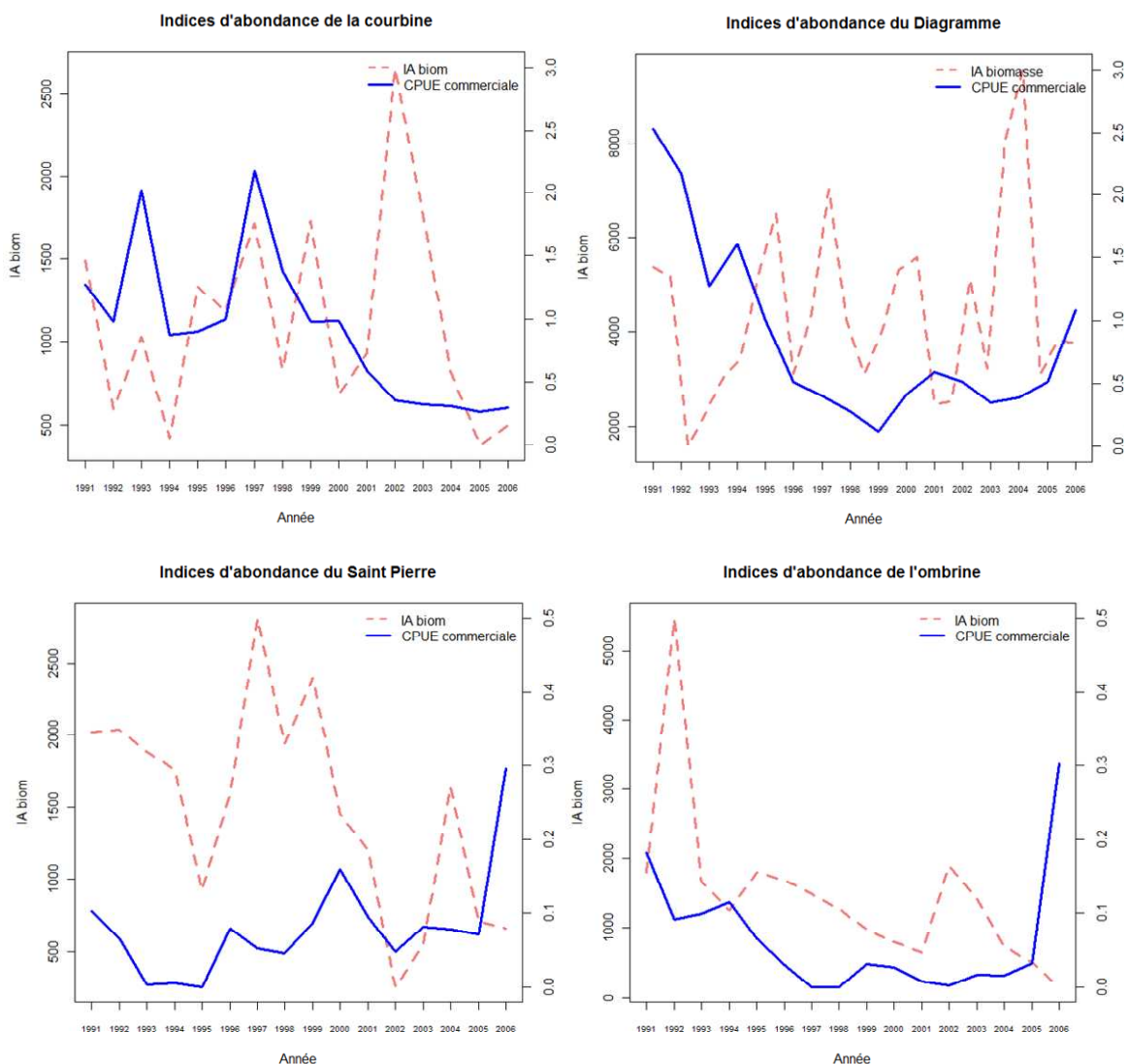


Figure 5 : Indices d'abondances standardisés de quatre poissons démersaux pêchés au large des côtes mauritaniennes.

### 3.3. Diagramme (*Plectorinchus mediterraneus*)

L'indice d'abondance issu des CPUE baisse depuis le début de la période jusqu'en 1999 puis semble remonter légèrement avec des fluctuations depuis lors. Ceci peut être liée soit à un changement de stratégie des navires nationaux dans leur ciblage suite à la rareté des céphalopodes. Ou bien ceci est le reflet réel de l'abondance du Diagramme. Cette dernière explication peut être crédible étant donné que cette espèce à affinité saharienne rencontrée sur toute la côte mauritannienne se nourrit essentiellement de crustacés (fischer et al., 1981), qui se trouvent eux même en forte expansion ces dernières années dans les eaux mauritaniennes. Par contre l'indice des campagnes scientifiques fluctue sans qu'une nette tendance de baisse ou d'accroissement soit mise en évidence.

Cependant un pic trois fois plus important que la moyenne sur toute la période est observé en 2003. Cette divergence des séries d'abondances devrait nous conduire à nous intéresser aux différents diagnostics qui découleront de l'ajustement du modèle global sur ces différents indices.

### 3.4. Saint Pierre (*Zeus faber*)

Les deux indices fluctuent avec des tendances globalement opposées (Figure 5). Tandis que l'indice d'abondance semble en légère baisse dans les campagnes, il est en légère hausse d'après les CPUE. Cette hausse pourrait être due d'un ciblage croissant du Saint Pierre par les flottilles industrielles. Notons qu'avant 1998 les deux indices semblent être corrélés.

### 3.5. Ombrine (*Umbrina canariensis*)

Une nette tendance à la baisse de l'abondance de l'ombrine est observée sur toute la période dans les campagnes scientifiques et dans les CPUE des navires nationaux, exception faite pour la dernière année (Figure 5).

## 4. DISCUSSION & CONCLUSIONS

### 4.1. Limites et sources de biais des CPUE comme indices d'abondances

La première source de biais est la possible non proportionnalité entre les CPUE et l'abondance réelle (une des conditions de la proportionnalité est l'hypothèse d'une distribution aléatoire de l'effort de pêche, hypothèse rarement vérifiée ; Hilborn et Walters 1992). Il est notamment difficile d'intégrer dans le temps l'évolution des puissances de pêche des navires liée soit à des processus d'innovations technologiques, soit à des changements de répartition spatiale de leur effort de pêche. En particulier, le changement des rendements au cours des années peut être lié au changement des stratégies de pêche des bateaux commerciaux, suite aux contraintes de marchés ou bien à la raréfaction d'une espèce cible comme le poulpe chez les Céphalopodiens nationaux mauritaniens.

Dans notre cas, l'effet "zone de pêche" n'est pas pris en compte car cette information manque dans les débarquements qui sont agrégés par marées. Même si on peut connaître la répartition par zone des captures totales et de l'effort de pêche fourni au sein de chaque marée à partir de la base journaux de bord, il reste impossible de savoir quelle espèce a été pêchée dans quel secteur. Une solution consisterait à ventiler les captures spécifiques par secteur selon l'importance des captures agrégées dans chaque secteur ; cela n'a pas été fait dans la présente étude.

### 4.2. Limites des indices issus des campagnes scientifiques

Le principal reproche qui peut être fait aux indices de campagnes scientifiques est le faible taux d'échantillonnage, notamment le faible nombre de traits effectués par strate (bathymétrie ou zone) suite au coût très élevé de ces campagnes. A titre d'exemple un jour de mer coûte en moyenne l'équivalent de 4 000 euros à l'IMROP, avec une réalisation de 5 à 10 traits de pêche (d'une durée de 30 minutes) par jour. La campagne dure une quinzaine de jours, et couvre une surface de l'ordre de 5 km<sup>2</sup>, alors que le plateau continental est d'une superficie de 33 000 km<sup>2</sup> environ.

### 4.3. Conclusions

Les évolutions des indices d'abondances des différentes espèces présentées montrent de fortes fluctuations sur la période étudiée.

Globalement les modèles expliquent peu de variabilité pour la majorité des espèces. Ceci peut être lié à la non prise en compte de certains facteurs, notamment la zone de pêche qui ne peut être testée vu que les captures sont agrégées par marée. On peut aussi penser à d'autres facteurs comme l'expérience des capitaines et la performance des navires en terme d'équipement de pêche.

Globalement, les évolutions d'abondance sont contrastées. Certaines espèces sont en fort déclin sur toute la période, comme le thiof, et durant les dernières années comme la courbine. A l'inverse d'autres semblent être légèrement en diminution à savoir l'Ombrine, voir même en augmentation comme le Saint Pierre d'après les CPUE.

Pour certaines espèces les tendances traduites par les CPUE commerciales semblent être plus cohérentes avec la connaissance issue de l'expertise, que les résultats issus des campagnes scientifiques. Par exemple, la nette diminution de l'abondance de la Courbine depuis 1997 est constatée sur le terrain, tandis que l'indice des campagnes scientifique ne fait apparaître cette chute que tout récemment. Ceci peut être expliqué par le fait que le navire de recherche a du mal à capturer cette espèce sémi-pélagique, même en période de forte abondance. A l'inverse, les navires commerciaux sont plus enclin à capturer cette espèce, vu l'importance du temps de pêche fourni.

A l'inverse, pour certaines espèces pour lesquelles l'expertise est moins claire, on accordera plus de crédit aux indices standardisés des campagnes scientifiques. Cela concerne notamment le diagramme, l'ombrine et le Saint Pierre.

La saisonnalité des captures est très significative chez la Courbine, qui est moins présente en juillet et août, période qui correspond au début de la saison chaude en Mauritanie. Tixerant (1974) et Limouzy (1983) avait signalé qu'en Mauritanie pendant la saison chaude les courbines migrent vers le Nord pour trouver des eaux plus froides, puis entrent dans la baie du lévrier<sup>1</sup> (en Mauritanie) à nouveau pendant les premiers mois de la saison froide, avant de continuer plus au sud. Notons qu'à l'échelle mondiale, trois sites de reproduction sont connus pour cette espèce à savoir : l'estuaire du Nil en Egypte, l'estuaire de la Gironde en France et la baie du lévrier en Mauritanie où elle est présente toute l'année (Quéméner, 2002).

Par contre, chez le thiof aucune saisonnalité n'est mise en évidence. Pourtant plusieurs auteurs (Domaine, 1980 ; Cury et al., 1982 ; in Dah A., Girardin M., M. vall, 1991) ont noté son comportement migratoire saisonnier, comme pour la majorité des espèces de la communauté à Sparidés à affinités sahariennes. Par contre, ces mêmes auteurs précisent que seuls les adultes accomplissent des migrations massives et que les jeunes sont plus sédentaires. Donc l'absence d'effet mois dans notre cas d'étude s'expliquerait par la présence régulière d'individus dans les captures quelque soit leur taille (l'effet saisonnalité est masqué par les petites tailles). La prise en compte des tailles de poissons débarqués confirmerait cette hypothèse.

Ces variations paraissent refléter au moins autant des changements dans le choix des espèces cibles que dans l'abondance absolue des espèces. Ces changements de stratégie dans l'exploitation des poissons (y compris dans les taux de rejet) réduisent beaucoup le niveau de significativité des indices d'abondance et des efforts effectifs tirés des statistiques commerciales (M. Fall et D. Leung Tack, GT 1991). Ceci justifie également la plus grande confiance accordée aux indices issus des campagnes scientifiques.

---

<sup>1</sup> La baie du Lévrier à Nouadhibou constitue en fait un habitat idéal pour cette espèce en raison des eaux peu profondes, de ses fonds de vase et de l'abondance de nourriture qui s'y trouve. Des chalutages expérimentaux effectués au milieu de la baie ont confirmé la présence de juvéniles sur les fonds (10 à 17 m), tout au long de l'année, ayant des températures de tolérances et des préférences différentes des adultes dans la baie (Limouzy, 1978).

## 5. ANNEXE : SCRIPT R DU MODELE GLM (CAS DU THIOF)

```
# download the lattice Package (to improve graphics)
library(lattice)
# Download other graphic functions
source("fonctionerror.bar.R")
source("gamhist.R")

# Read the data
ALLCPUE <- read.csv("Max CPUE avec zéros.csv",header=TRUE, sep=";")
ALLCPUE[1:5,]
Campagne <- read.csv("GLM campagnes.csv",header=TRUE, sep=";")
summary(ALLCPUE)
summary(Campagne)

# Retrait de l'année 2007
ALLCPUE2<-ALLCPUE[-which(ALLCPUE$Année==2007),]
unique(ALLCPUE2$etat1)

# On impose le système de contraintes "somme = 0" sur les facteurs
contrasts(ALLCPUE2$Année) <- contr.sum(levels(ALLCPUE2$Année))
contrasts(ALLCPUE2$ANCONS) <- contr.sum(levels(ALLCPUE2$ANCONS))
contrasts(ALLCPUE2$TYPNAV) <- contr.sum(levels(ALLCPUE2$TYPNAV))
contrasts(ALLCPUE2$TJBclas1) <- contr.sum(levels(ALLCPUE2$TJBclas1))
contrasts(ALLCPUE2$TYPNAV) <- contr.sum(levels(ALLCPUE2$TYPNAV))
contrasts(ALLCPUE2$Mois) <- contr.sum(levels(ALLCPUE2$Mois))
contrasts(ALLCPUE2$etat1) <- contr.sum(levels(ALLCPUE2$etat1))

# Modèle de presence / Absence (GLM Binomial)
# -----
THIOFbin.glm <- glm(PA.THIOF ~ Année + TJBclas1 + Année:Mois, family = binomial, data =
ALLCPUE2)

# model output
```

```
summary(THIOFbin.glm)
anova(THIOFbin.glm,test="Chisq")

# Log-Gaussian model on positive values
# -----
# Extract positive values
thiofpresent <- subset(x = ALLCPUE2, subset = ALLCPUE2$THIOF>0) ;
head (thiofpresent$PA.THIOF)

THIOFpresent.glm <- glm(log(THIOF) ~ TJBclas1 + Année + TYPNAV + Année + Mois, family =
gaussian, data = thiofpresent)

# Model output
summary(THIOFpresent.glm)
anova(THIOFpresent.glm,test="Chisq")

# Diagnostics graphiques
windows()
par(mfcol=c(2,2))
plot(THIOFpresent.glm)
par(mfcol=c(1,1))

# Predictions
# -----

# Création de la matrice de croisement des facteurs pour les prédictions

L_An <- length(levels(as.factor(ALLCPUE2$Année)))
L_TJB <- length(levels(ALLCPUE2$TJBclas1))
L_Mois <- length(levels(ALLCPUE2$Mois))
L_TYPNAV <- length(levels(ALLCPUE2$TYPNAV))
L_tot <- L_An* L_TJB* L_Mois* L_TYPNAV
Table_Pred <- matrix(NA,nrow=L_tot,ncol=4)
#col 4
```

```

Table_Pred[,4] <- rep(levels(ALLCPUE2$TYPNAV),(L_tot/L_TYPNAV))
#col 3 , col 2 & col 1
for(k in 1:L_An){
  for(j in 1:L_TJB){
    for(m in 1:L_Mois){
      deb_An <- k*L_TJB*L_Mois*L_TYPNAV - L_TJB*L_Mois*L_TYPNAV
      fin_An <- k*L_TJB*L_Mois*L_TYPNAV
      deb_TJB <- deb_An + j*L_Mois*L_TYPNAV - L_Mois*L_TYPNAV
      fin_TJB <- deb_An + j*L_Mois*L_TYPNAV
      deb_Mois <- deb_TJB + m*L_TYPNAV - L_TYPNAV+1
      fin_Mois <- deb_TJB + m*L_TYPNAV
      Table_Pred[deb_Mois:fin_Mois,3] <- rep(levels(ALLCPUE2$Mois)[m], L_TYPNAV)
    }
    Table_Pred[(deb_TJB+1):fin_TJB,2] <- rep(levels(ALLCPUE2$TJBclas1)[j], L_Mois*L_TYPNAV)
  }
  Table_Pred[(deb_An+1):fin_An,1] <- rep(levels(as.factor(ALLCPUE2$Année))[k], L_TJB * L_Mois
  * L_TYPNAV)
}
Table_Pred[1:20,]

results <- data.frame(Table_Pred)
names(results) <- c("Année", "TJBclas1", "Mois","TYPNAV")
results[1:10,]
results.thioff <-results[-which(results$Année %in% c(2006,2007)),]

# Prédiction des probabilités de présence
#test logique
toto <- if(ALLCPUE2$PA.THIOF!=0){ "vrai" } else {"faux"}
predictionspresence <- predict.glm(THIOFbin.glm, results.thioff, type="r", se=T)
predictionspresence$fit[1:10]

# Prediction densités positives en échelle log
predictionscpue < -predict.glm(THIOFpresent.glm,results.thioff,type="r",se=T)

```

```
predictionscpue$fit[1:10]
```

```
# Prédiction finale en échelle LOG densité
```

```
# -----
```

```
predictionVRAICPUE<-results
```

```
predictionVRAICPUEfit <- predictionspresence$fit*predictionscpue$fit
```

```
# Prédiction finale dans l'échelle de la densité avec transformation LOG inverse
```

```
# -----
```

```
varlogthiofOposit <- (predictionscpue$se.fit)^2
```

```
predictionVRAICPUEfit <- predictionspresence$fit * exp(predictionscpue$fit + 0.5*  
varlogthiofOposit)
```

```
##### GRAPHS
```

```
CPUEthiof<-presencethiof0*thiofOposit
```

```
test<-data.frame(CPUEthiof<-presencethiof0*thiofOposit)
```

```
head(test)
```

```
#le graph de la CPUE standardisée par GLM
```

```
windows()
```

```
par(mfcol=c(1,1))
```

```
plot(CPUEthiof,pch=1, type="l", lty=3, main="CPUE/GLM du THIOF", col="white")
```

```
lines(CPUEthiof,lty=1,lwd=2,col="blue")
```

```
par(mfcol=c(1,1), col="blue")
```

```
#Probabilités de présence par année
```

```
head(presencethiof0)
```

```
test3<-data.frame(presencethiof0)
```

```
head(test3)
```

```
test4<-cbind(results.thioff,test3)
```

```
#concataner des colonnes
```

```
head(test4)
names(test4) <- c("Année", "TJBclas1", "Mois", "TYPNAV", "PA")
head(test4)
dim(test4)
PAmoythiof<-aggregate(test4$PA,list(Année=test2$Année),FUN=mean)
PAmoythiof

#le graph des probabilités de présence
windows()
par(mfcol=c(1,1))
plot(PAmoythiof,pch=2, type="l", main="Probabilités de présence du THIOF dans les marées",
col="red")
lines(PAmoythiof, type="l")
par(mfcol=c(1,1), col="blue")

##CPUE non standardisée pour les valeurs positives uniquement
thiofNSposit<-read.csv("thiof.csv",header=TRUE, sep=";")
thiofNSposit[,2] <- as.factor(thiofNSposit[,2])
head(thiofNSposit)
CPUENSthiof<-
aggregate(thiofNSposit$CPUeHeures,list(Année=thiofNSposit$Année),FUN=mean)
CPUENSthiof

windows()
par(mfcol=c(1,1))
plot(CPUENSthiof,pch=1, type="l", main="CPUE NON STD du THIOF (MRT)", col="red")
lines(CPUENSthiof, lty=2, lwd=1)

library(lattice)

###graphe superposé
windows()
par(mfcol=c(1,1))
plot(x=CPUENSthiof$Année, y=CPUENSthiof$x, type="l",
      lwd=1,lty=2, col="black", main="CPUE/GLM du THIOF")
```



```

lines(x=CPUETHIOF$ Année[1:15],y=CPUETHIOF$ CPUE,
      pch=1, type="l", lwd=3,lty=1, col="blue")

windows()
par(mfcol=c(1,1))
plot(x=as.numeric(levels(CPUENSTHIOF$ Année)), y=CPUENSTHIOF$x, type="l",
      lwd=1,lty=2, col="black", main="CPUE du THIOF")
lines(x=as.numeric(levels(CPUETHIOF$ Année)[1:15]),y=CPUETHIOF$ CPUE,
      pch=1, type="l", lwd=3,lty=1, col="blue")
#legende(legende=c("année","cpue")
#legende(text.legend,xlab=" Année", ylab="CPUE",x="topleft", col("red,"blue"), title="légende")

#graphes superposés avec campagnes
campagne<-read.csv("GLM campagnes.csv",header=TRUE, sep=";")
campagne
CPUETHIOF

windows()
par(mfcol=c(1,1))
plot(x=as.numeric(campagne$ Année), y=campagne$ Thiof, type="l",
      lwd=3,lty=2, col="LightCoral", main="Indices d'abondance de THIOF", ylim = c(0,3),
      xlab=" Année",ylab="",xaxt="n")
lines(x=as.numeric(levels(CPUETHIOF$ Année)[1:15]),y=CPUETHIOF$ CPUE,
      pch=1, type="l", lwd=3,lty=1, col="blue")

axis(1,at=seq(from=1991, to = 2006, by = 1), cex.axis = 0.7)
#xaxt pour enlever la première affiche des années

#Pour afficher les années en vertical
legend(legend=c("IA campagnes","CPUE commerciale"), x="topright", bty="n",lwd = 2:3, lty = 2:1
,col = c("LightCoral","blue"))
#legende(text.legend,xlab=" Année", ylab="CPUE",x="topleft", col("red,"blue"), title="légende")

windows()

```

```
plot(x=as.numeric(biom$an),y=biom$Epinephelus.aeneus,
pch=1, type="l", lwd=3,lty=1, col="blue", xlab="",ylab="", axes=FALSE)
```

```
axis(1,at=seq(from=1982, to = 2006, by = 1), cex.axis = 0.7)
```

```
axis(4)
```

```
#xaxt pour enlever la première affiche des années
```

```
windows()
```

```
par(new=TRUE)
```

```
plot(x=as.numeric(biom$an),y=biom$Epinephelus.aeneus,
pch=17, type="l", lwd=3,lty=4, col="blue", xlab="",ylab="", axes=FALSE)
```

```
par(new=TRUE)
```

```
plot(x=c(1982:1990,as.numeric(levels(thiofcpue$ Année)[1:16 ])),y=c(rep(NA,9),thiofcpue$thiof),
type="l", lwd=3, lty=1, col="blue",ylim=c(0,3), xlab="",ylab="", axes=F)
```

```
axis(1,at=seq(from=1982, to = 2006, by = 1))
```

```
axis(4)
```

```
#xaxt pour enlever la première affiche des années
```

```
par(new=TRUE)
```

```
plot(x=as.numeric(campagne$ Année), y=campagne$ Thiof, type="l",
```

```
lwd=3,lty=2, col="LightCoral", main="Indices d'abondance du thiof", xlab="Année",ylab="IA
biom",xaxt="n")
```

```
#pour afficher les années en vertical
```

```
legend(legend=c("IA rend", "IA biomas", "CPUE"), x="topright", bty="n",lwd = 2:3, lty = 2:1 ,col =
c("LightCoral", "blue", "brown"))
```

```
#legend(text.legend,xlab=" Année", ylab="CPUE",x="topleft", col("red","blue","brown"),
title="légende")
```

```
windows()
```

```
par(mfcol=c(1,1))
```

```
plot(x=as.numeric(campagne$ Année), y=campagne$ Thiof, type="l",
```

```
lwd=3,lty=2, col="LightCoral", main="Indices d'abondance du thiof", xlab="Année",ylab="IA  
biom",xaxt="n")
```

```
par(new=TRUE)
```

```
plot(x=as.numeric(biom$an),y=biom$Epinephelus.aeneus,  
pch=1, type="l", lwd=3,lty=1, col="blue",ylim=c(0,3), xlab="",ylab="", axes=FALSE)  
axis(1,at=seq(from=1991, to = 2006, by = 1))  
axis(4)
```

```
plot(x=as.numeric(thiofcpue$Année),y=thiofcpue$thiof,  
pch=1, type="l", lwd=3,lty=1, col="blue",ylim=c(0,3), xlab="",ylab="", axes=FALSE)
```

```
axis(1,at=seq(from=1991, to = 2006, by = 1), cex.axis = 0.7)  
axis(4)
```

```
#xaxt pour enlever la première affiche des années
```