

Modélisation statistique

Marie-Pierre Etienne

<https://github.com/MarieEtienne>

Septembre 2018



Outline

- 1 Introduction
- 2 Démarche de modélisation

Outline

① Introduction

② Démarche de modélisation

Motivations

- **Comprendre** : des mécanismes biologiques, économiques, ...

- **Comprendre** : des mécanismes biologiques, économiques, ...



Habitat suitability for juvenile common sole (*Solea solea*, L.) in the Bay of Biscay (France): A quantitative description using indicators based on epibenthic fauna

Olivier Le Pape^{a,*}, Loïc Baulier^a, Aurélie Cloarec^a, Jocelyne Martin^b, François Le Loc'h^c, Yves Désauvay^b

^a Département Halieutique, Agroparcus Rennes, 65 rue de St Brice, CS 84215, 35042 Rennes Cedex, France

^b IFREMER, 33400, rue de l'Île d'Yeu, B.P. 21103, 44111 Nantes Cedex, France

^c IRD, UR RAP, Centre de Recherche Halieutique, Avenue Jean Monnet, B.P. 171, 34203 Sète Cedex, France

Received 28 September 2005; accepted 31 August 2006

Available online 16 September 2006

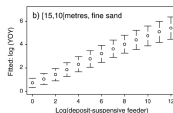


Fig. 3. Fitted log-transformed young-of-the-year common sole *Solea solea* abundance (calculated by delta model with Eq. (6), built with Eqs. (1) and (5) and associated confidence intervals (calculated with Eq. (7), for $\alpha=0.05$) in different physical habitats, with relation to the density of trawled epibenthic invertebrates that feed on detrital or planktonic organic matter.

Type de modèles :

$$\ln(\text{YOY}_+) \approx (\text{nursery} \times \text{bathymetry} \times \text{sediment}) + \text{covariate}(\ln(\text{func.group} + 1)) + \varepsilon_+ \quad (4)$$

or

$$\ln(\text{YOY}_+) \approx (\text{bathymetry} \times \text{sediment}) + \text{covariate}(\ln(\text{func.group} + 1)) + \varepsilon_+ \quad (5)$$

Motivations

- **Comprendre** : des mécanismes biologiques, économiques, ...
- **Prédire** : des tendances, des choix,

- **Comprendre** : des mécanismes biologiques, économiques, ...
- **Prédire** : des tendances, des choix,

IOTC-2016-WPTT18-19
Received: 1 November 2016

Bayesian state-space production models for the Indian Ocean bigeye tuna (*Thunnus Obesus*) and their predictive evaluation

Kento Otsuyama and Toshihide Kitakado*

¹Tokyo University of Marine Science and Technology, 5-7, Kanari, Minato, Tokyo, 108-8477, Japan

*Corresponding author: kitakado@kaiyodai.ac.jp

Abstract

In stock assessment, it is not straightforward to choose a plausible range of models objectively from several models if different data set are used because it is not possible to use model selection criteria like AIC in these situation. However, as shown in Kell et al. (2016), where a hindcasting approach was proposed, predictive evaluation via cross-validation would be a possible procedure under those circumstances. Here, as an attempt using data for bigeye tuna, we applied a model selection method with predictive evaluation of biomass index to Bayesian state-space production models although the data used is common to the model in this case. Using a selected model, we also assessed the population status of the stock. Non-informative priors were used and posterior samples were generated using a Markov chain Monte Carlo (MCMC) method. The results suggested that \hat{F} -ratio (2015) is higher than the MSY level (1.17) and \hat{B} -ratio is lower than 1 (0.76). Given that this analysis has a preliminary nature as stock assessment, the paper may not be so useful for management advice, but this approach could give an opportunity to help in choosing models in future assessment.

Motivations

- **Comprendre** : des mécanismes biologiques, économiques, ...
- **Prédire** : des tendances, des choix,
- **Communiquer** : sur l'état d'un système, les conséquences d'une décision

SCRS/2015/168

Collect. Vol. Sci. Pap. ICCAT, 72(2): 564-576 (2016)

PROJECTIONS, KOBE PLOTS, AND MAXIMUM SUSTAINABLE YIELDS FOR ATLANTIC BIGEYE TUNA 2015

Michael J. Schirripa¹

SUMMARY

This work is intended to accompany the 2015 bigeye tuna assessment by providing projections, KOBE plots and estimates of maximum sustainable yield for the work done within the Stock Synthesis framework. The combined model estimates the probability of fishing mortality in 2015 being greater than F_{MSY} of 80% and biomass being less than B_{MSY} of 99 percent. If the current catches of approximately 70,000 mt are continued the stock is not projected to rebuild to $B/B_{MSY} > 1$ with at least a 50% probability by 2028. To rebuild to $B/B_{MSY} > 1$ with at least 50% probability by 2026 (i.e. ten years) would require a total allowable catch of no greater than 60,000 mt. The constant catch projections of B/B_{MSY} and F/F_{MSY} provided assume that the proportion of catch from each fleet will remain constant within the future time frame considered (2015-2028). However, this assumption is not well supported by the historic trends. As the fishery continues to focus on increasing younger fish the year specific estimates of MSY decrease while the stock size required to achieve the lower MSY increases.

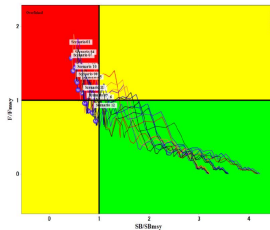
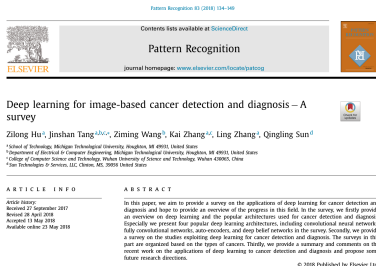


Figure 1. Kobe phase plot for all twelve SS scenarios considered.

Les outils : différents types de modèles

- **Machine Learning** (deep learning)



Modèles boîtes noires, gourmands en données avec un but purement prédictif (réseaux de neurones, forêts aléatoires, ...)

- **Modèle stochastique**

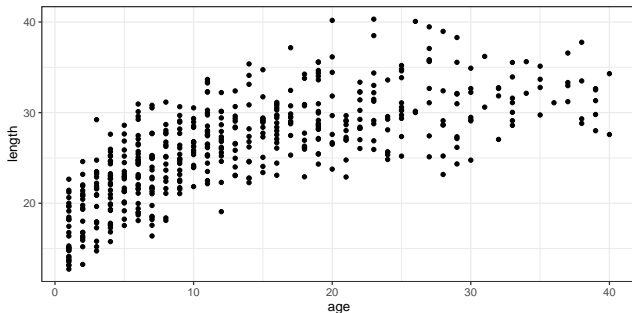
De la simple mesure de corrélation jusqu'au modèle mécaniste ...

- **Modèle déterministe**

Le plus souvent dans la panoplie des modèles mécanistes

Un exemple : relation von Bertalanffy

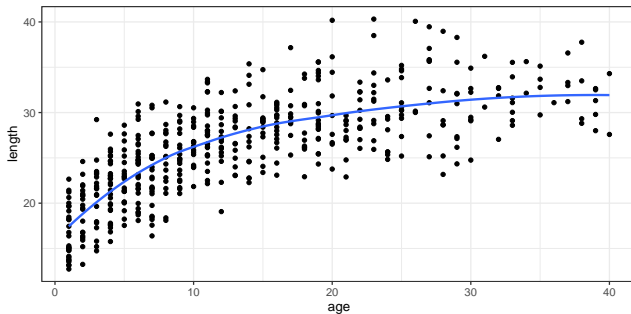
Des données de croissance



Un exemple : relation longueur et âge

Des données de croissance

Ajustement, 'lissage' par la méthode 'loess'

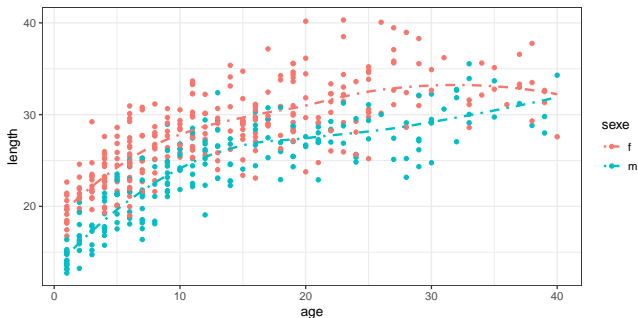


On veut représenter le comportement moyen.

Un exemple : relation longueur et âge

Des données de croissance

Ajustement, 'lissage' par la méthode 'loess'

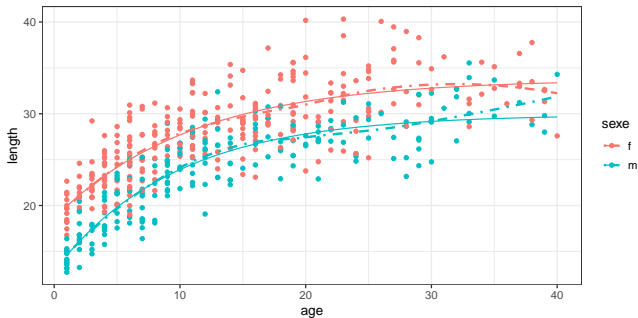


On ne peut pas facilement tester si les deux courbes sont significativement différentes.

Un exemple : relation longueur et âge

Ajustement d'une relation de von Bertalanffy

$$L_k = L_{\text{inf}} \left(1 - e^{\{-K(x-x_0)\}} \right)$$



On a perdu en ajustement mais on peut tester une différence mâle femelle.

Un exemple : relation von Bertalanffy

```
summary(estim_nls)
```

```
##
## Formula: length ~ (ldiff_m * sexe_num + Linf) * (1 - exp(-(kdiff_m * sexe_num +
##      K) * (age - (xdiff_m * sexe_num + x0))))
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## Linf      33.76311    0.64015  52.743 < 2e-16 ***
## K          0.09158    0.01168   7.838 2.84e-14 ***
## x0        -8.73180    1.28783  -6.780 3.43e-11 ***
## ldiff_m   -3.85000    0.88394  -4.355 1.62e-05 ***
## kdiff_m    0.01269    0.01690   0.751  0.4529
## xdiff_m    3.32673    1.54949   2.147  0.0323 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.752 on 494 degrees of freedom
##
## Number of iterations to convergence: 10
## Achieved convergence tolerance: 1.17e-06
```

Un exemple : relation von Bertalanffy

14

Equation pour représenter un processus biologique/écologique

Modèle de croissance von Bertalanffy

$$\frac{dw}{dt} = \text{Anabolisme} - \text{Catabolisme}$$

Équation de flux pour un système ouvert

$$\frac{dw}{dt} = Hs - kw$$

H : taux de synthèse / u. surface d'absorption ;

k : taux de dégradation / u. poids

$$\frac{dL^3}{dt} = H'L^2 - k'L^3 \Rightarrow 3L^2 \frac{dL}{dt} = H'L^2 - k'L^3 \Rightarrow \frac{dL}{dt} = H' - k'L$$

$$\Rightarrow \left\{ \begin{array}{l} L(t) = L_{\infty} - (L_{\infty} - L_0) e^{-kt} \\ L(t) = L_{\infty} (1 - e^{-k(t-t_0)}) \\ w(t) = w_{\infty} (1 - e^{-k(t-t_0)})^3 \end{array} \right.$$

t₀ : Temps au quel L = 0 et w = 0;
Une constante artificielle, obtenue par ajustement sur des données

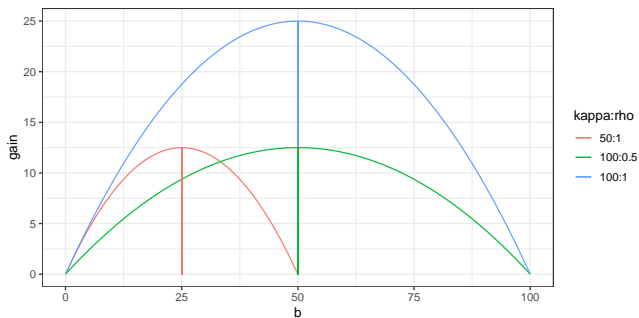
Source : Raymond J.H. Beverton and Sidney Holt. On the dynamics of exploited fish populations.

D'après L. Von Bertalanffy, 1934, 1938, 1949

Modèle de production de biomasse

$$B_{t+1} = B_t + g(B_t - C_t), \quad g(b) = \rho b \left(1 - \frac{b}{\kappa}\right)$$

$$B_{MSY} = \frac{\kappa}{2}, \quad C_{MSY} = \frac{\kappa\rho}{4}$$



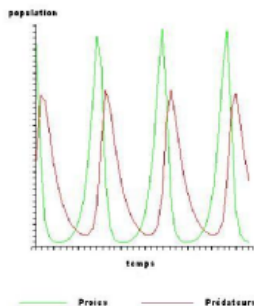
Dynamique Proies-Prédateurs

■ Lotka-Volterra

x_t **Densité proies**

y_t **Densité prédateur**

$$\frac{1}{x_t} \frac{dx_t}{dt} = \underbrace{a - b \cdot x_t}_{\text{Internal competition}} - \underbrace{c \cdot y_t}_{\text{Predation}}$$
$$\frac{1}{y_t} \frac{dy_t}{dt} = \underbrace{d - e \cdot y_t}_{\text{Internal competition}} + \underbrace{f \cdot x_t}_{\text{Predation}}$$



Modèle de réseaux trophiques

www.ecopath.org

Modéliser les flux
(matière, énergie)
dans un réseau
trophique
(n compartiments)

Flux à l'état stable

Production =
Capture +
Consommation +
Mortalité +
Transferts

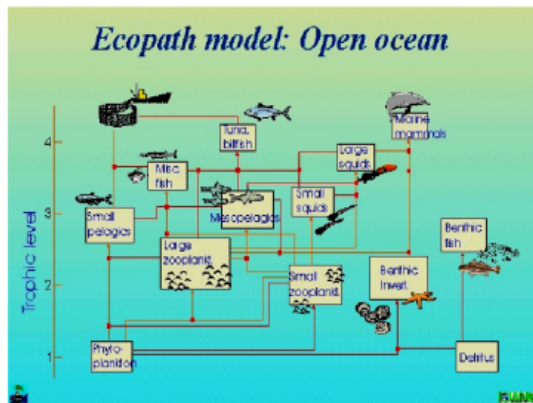


Figure 3: Etienne Rivot : Ecopath

Modèles « End-to-end » - Couplage IBM / Modèles NPZ

Modèle IBM (multispécifique)

Modèle N - P - Z

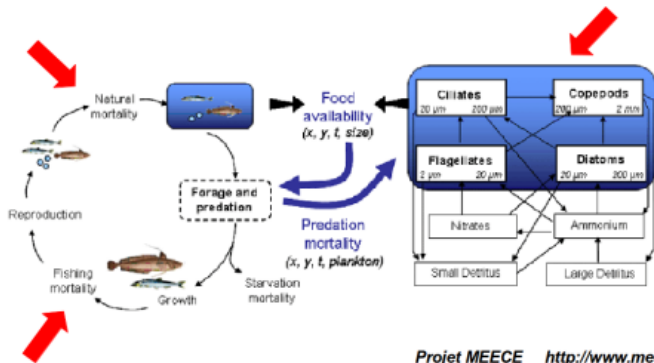


Figure 4: Etienne Rivot : IBM

Outline

① Introduction

② Démarche de modélisation

Comment se lancer

Il n'existe pas un unique bon modèle qui répond à toutes les questions

—

Douglas Adams Le Guide du voyageur galactique :

Quarante-deux ! cria Loonquawl. Et c'est tout ce que t'as à nous montrer au bout de sept millions et demi d'années de boulot ?

J'ai vérifié très soigneusement, dit l'ordinateur, et c'est incontestablement la réponse exacte. Je crois que le problème, pour être tout à fait franc avec vous, est que vous n'avez jamais vraiment bien saisi la question.

—

Essentially, all models are wrong, but some are useful. (Box, George E. P.; Norman R. Draper)

Comment se lancer

- ① Définir l'objectif de la modélisation : prédire, comprendre,
- ② Adopter un principe de parcimonie : choisir le plus petit niveau de complexité raisonnable pour répondre à la question.

Le rôle des données

- Pour émettre des hypothèses
- Pour ajuster le modèle
- Pour tester les qualités du modèle

Comment s'y prendre

- une démarche itérative : on ne construit jamais un modèle en une fois.
- des outils pour estimer les paramètres : MCO, max de vraisemblance
- des outils pour vérifier que le modèle est identifiable (au moins en pratique)
- des outils pour juger les qualités prédictives d'un modèle

Connaître un peu de théorie

« He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may be cast »

Leonardo Da Vinci

« Celui qui veut faire correctement son travail doit d'abord aiguïser ces outils »



Confucius - Analectes (les Entretiens)
(Chine, vers 550 av JC)

Figure 5: Etienne Rivot : conclusions