

Problématiques de calibration, analyse de
sensibilité et classification d'un modèle DEB de
Yellowfin

IRD-IMT, 26 mai 2015

Un essai de formalisation mathématique du contexte

- ▶ Soit $L(t)$ la longueur d'un animal à l'âge t . Cette courbe est croissante avec t et bornée pour des raisons biologiques
- ▶ Pour un ensemble de pas de temps fixés, on dispose d'un modèle de simulation

$$L(t) = g_{\theta}(X)$$

où :

1. $X \in \Omega \subset \mathbb{R}^d$ est un vecteur de paramètres environnementaux et individus-centrés
 2. $\theta \in \Theta \subset \mathbb{R}^q$ est un vecteur de paramètres dont la valeur est intrinsèque à un couple (individu, espèce), par ex. ceux d'un modèle de métabolisme DEB
- ▶ \Rightarrow la loi $f(x)$ représente donc la distribution des environnements possibles d'un animal en particulier
 - ▶ \Rightarrow la loi $\pi(\theta)$ dirige la croissance type d'un individu de l'espèce considérée

1. Raffinement des lois d'entrée $f(x)$ et $\pi(\theta)$ par l'utilisation d'observations de longueur idéalement telles que

$$L^*(t_i) = L(t_i|x, \theta) + \xi_i \text{ avec } \xi_i \sim \mathcal{N}(0, \sigma^2), \quad (1.1)$$

2. Expliquer pour quels X et θ (quelles lois conditionnelles) on peut obtenir des courbes de croissance dont la dynamique diffère, tout en restant pertinentes biologiquement ; en particulier, peut-on discriminer des conditions environnementales (sur X) menant à des courbes possédant une, deux ou trois stances ?
3. Influence des variables ? Nécessité d'une étude de sensibilité à maille de plus en plus fine
4. Quid de l'erreur de modèle sur $g_\theta(X)$?

Descriptif des *inputs* (1/2)

```
kappa <- param[i,1] # Fraction of mobilised reserve allocated to soma
pAm <- param[i,2] # Maximum assimilation rate (J/d/cm^2)
pMp <- param[i,3] # 1st volume-specific somatic maintenance rate (J/cm^3/d)
pMj <- param[i,4] # 2nd volume-specific somatic maintenance rate (J/cm^3/d)
pT <- param[i,5] # Surface area-specific somatic maintenance rate (J/cm^2/d)
v <- param[i,6] # Energy conductance (cm/d)
EG <- param[i,7] # Volume-specific cost of structure (J/cm^3)
E0 <- param[i,8] # Initial energy reserve (energy costs of one egg J)
d1 <- param[i,9] # Shape parameter # 1 (larval stage)
d2 <- param[i,10] # Shape parameter # 2 (adult stage)
dv <- param[i,11] # Structural volume density (g/cm^3)
rho_E <- param[i,12] # Weight-energy coupler (J/g)
kj <- param[i,13] # Maturity maintenance rate coefficient (d-1)
Xk <- param[i,14] # Half saturation constant
TA <- param[i,15] # Arrhenius temperature (K)
T1 <- param[i,16] # Reference temperature (K)
EHp <- param[i,17] # Maturation level at puberty threshold (J)
EHb <- param[i,18] # Maturation level at birth (J)
EHj <- param[i,19] # Maturation level at the end of the larval stage (J)
EHy <- param[i,20] # Maturation level at the end of the early juvenile phase (J)
Lp <- param[i,21] # Structural length at maturity (cm)
Lb <- param[i,22] # Structural length at birth (cm)
Lj <- param[i,23] # Structural volumetric length at metamorphosis (cm)
Lj1 <- param[i,24] # Structural volumetric length at end of early juvenile stage (cm)
kappa_R <- param[i,25] # Fraction of reproduction energy fixed in eggs
N <- param[i,26] # Mean egg number by batch
N_min <- param[i,27] # Minimal egg number by batch
Eb <- param[i,28] # Energy density for 1 batch (J.cm-3)
Ar <- param[i,29] # Age at first spawning (d)
```

Descriptif des *inputs* (2/2)

```
# Initial state features
#-----
#
# td = Start date (birth)
# tf = End date (death of the individual)
# step = Model time step (d)
#
#
# Environmental input features (X)
#-----
#
# T = Temperature (kelvin; boundaries tolerance: 273 - 400)
# X = Food
#
```

Lois *a priori* des *inputs*

- ▶ Quelle information sur $f(x)$ et $\pi(\theta)$?
- ▶ Valeurs nominales ?
- ▶ Bibliographie espèces proches ?
- ▶ Domaines bornés ?
- ▶ Valeurs fixées ? (dimension 29 pour θ)

Valeurs "nominales" utilisées par Emmanuelle Dortel

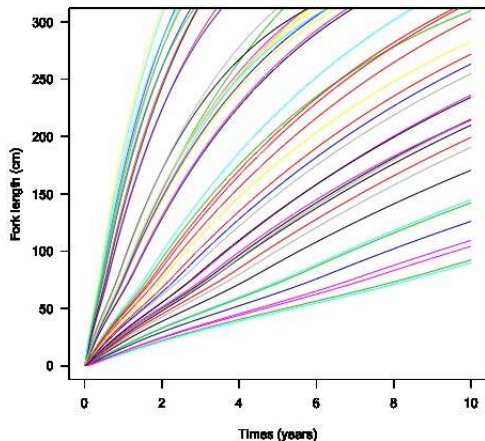
```
# Fraction of mobilised reserve allocated to soma
kappa <- 0.75
# Maximum assimilation rate (1/d/cm^2)
pAm <- 224 #J
# Volume-specific somatic maintenance rate (1/cm^3/d)
pMj <- 10.72
# Surface area-specific somatic maintenance rate (J/cm^2/d)
pMj <- 17.9 #J
# Surface area-specific somatic maintenance rate (J/cm^2/d)
pT <- 22.82 #J2
# Energy conductance (cm/d)
v <- 0.33 #J
# Volume-specific cost of structure (J/cm^3)
EG <- 8828 #J
# Initial energy reserve (energy costs of one egg J)
E0 <- 1.37921
# Shape parameter
d1 <- 0.2249 #J larval stage
d2 <- 0.2704 #J adulte stage
# Structural volume density (g/cm^3)
dv <- 1
# Weight-energy coupler (J/g)
rho_E <- 21146.5 #J
# Maturity maintenance rate coefficient (d-1)
kj <- 0.06303 #J
# Half saturation constant
Xk <- 0.000163
# Arrhenius temperature (K)
TA <- 9029
```

```
# Reference temperature (K)
T1 <- 298.65
# Maturation level at puberty threshold (J)
E1p <- 2548000 #J2
# Maturation level at birth (J)
E1b <- 0.7637 #J
# Maturation level at the end of the larval stage (J)
EHj <- 6902 #J
# Maturation level at the end of the early juvenile phase (J)
EHy <- 969500 #J
# Structural length at maturity (cm)
Lp <- 22.064 #J2
# Structural length at birth (cm)
Lb <- 0.0501778
# Structural volumetric length at metamorphosis (cm)
Lj <- 0.397276
# Structural volumetric length at end of early juvenile phase (cm)
Lj1 <- 10 #J2
# Fraction of reproduction energy fixed in eggs
kappa_R <- 0.95
# Mean egg number by batch
N <- 10000
# Minimal egg number by batch
N_min <- 5000
# Energy density for 1 batch (1.cm-3)
Eb <- 3.5
# Age at first spawning (d)
Ar <- 1131.448
```

50 réalisations par Monte Carlo

On fait varier uniformément entre 0.95 et 1.05 de la valeur nominale de chaque paramètre

Plantages code constatés (sans doute pour des entrées "impossibles")



Calibration par inversion : principe

Utiliser les données observées pour affiner la loi jointe *a priori* $f(x, \theta)$ sur chacun des domaines $\Omega_i \times \Theta_i$. Formellement, ce problème d'inversion consiste à estimer chaque loi

$$f_i(x, \theta | \mathbf{D}^*) \propto \ell(\mathbf{D}^* | x, \theta) f(x) \mathbb{1}_{\{x \in \Omega_i\}} \pi(\theta) \mathbb{1}_{\{\theta \in \Theta_i\}} \quad (1.2)$$

où $\ell(\mathbf{D}^* | x, \theta)$ est la vraisemblance des observations

L'atteinte de cette loi peut être réalisée au travers d'un algorithme fondé sur les chaînes de Markov, qui nécessite dès lors de très nombreux appels au modèle g

Approche par MCMC classique formellement codée

Approche par ABC (simulation progressive dans le prior $f(x, \theta)$) peut-être plus séduisante

⇒ comparaison des deux approches ?

Données de capture-recapture (RTTP-IO)

The in situ data \mathbf{D}^* are triplets $\{t_j^*, L^*(t_j), L^*(t_j + \Delta_j)\}_{j=1, \dots, n_2}$ corresponding to the observed age and to the lengths at capture and recapture, reconstituted using a reading model based on otolithometry (Dortel et al. 2013).

A difficulty comes from the fact that the real age t_i of first capture is unknown but estimated by t_i^* because of the reading error. It is assumed that

$$t_j^* = t_j + \nu_j \text{ with } \nu_j \sim \mathcal{N}(0, \tau^2). \quad (1.3)$$

and

$$L^*(t_j) = L(t_j|x, \theta) + \varepsilon_j^c \text{ with } \varepsilon_j^c \sim \mathcal{N}(0, \sigma_c^2), \quad (1.4)$$

$$L^*(t_j + \Delta_j) = L(t_j + \Delta_j|x, \theta) + \varepsilon_j^r \text{ with } \varepsilon_j^r \sim \mathcal{N}(0, \sigma_r^2). \quad (1.5)$$

Fork length measurements are obviously more precise for larger died fish than alive juveniles, hence $\sigma_r \leq \sigma_c$. Indeed, measures at first capture are made in stressing conditions (time < 10s) and cannot be repeated while the measures at recapture are conducted on frozen fish with callipers.

The likelihood of \mathbf{D}_2^* can be written under a conditional form:

$$\ell_2(x, \theta, \tau, \sigma_c, \sigma_r) = \prod_{j=1}^{n_2} [L^*(t_j + \Delta_j) | L^*(t_j), x, \theta, t_j^*, \tau, \sigma_c, \sigma_r] [L^*(t_j) | x, \theta, t_j^*, \tau]$$

where the symbol $[\cdot]$, popularized by Gelfand et al. (1990), indicates a (cumulative or density) probability distribution

Approximating the distribution of the length at capture

Consider the second term in (1.6), which corresponds to the law of a first observation (capture). One has

$$[L^*(t_j) | x, \theta, t_j^*, \tau, \sigma_c] = \int_0^\infty [L^*(t_j) | L(t_j | x, \theta), \sigma_c] [L(t_j | x, \theta) | t_j^*, \tau] d[L(t_j | x, \theta)].$$

The first integrand term is given by (1.8). An approximation of the second term can be produced by considering that

$$L(t_j|x, \theta) = L(t_j^*|x, \theta) + \sum_{k=1}^{\infty} \frac{(t_j^* - t_j)^k}{k!} L^{(k)}(t_j^*|x, \theta)$$

which exists by smooth regularity of the output at each time (age) step.

A first-order approximation gives (provided $L'(t_j^*|x, \theta) \neq 0$)

$$[L(t_j|x, \theta)|t_j^*, \tau] \sim \mathcal{N}\left(L(t_j^*|x, \theta), \tau^2 \{L'(t_j^*|x, \theta)\}^2\right)$$

and finally, benefiting from Gaussian conjugation and after some algebraic work, approximately

$$[L^*(t_j)|x, \theta, t_j^*, \tau, \sigma_c] \propto \frac{\mu_{c,t_j^*}}{\sigma_c \tau |L'(t_j^*|x, \theta)|} \exp\left(\frac{\eta_{c,t_j^*}^2}{2\mu_{c,t_j^*}^2} - \lambda_{c,t_j^*}^2\right)$$

where the symbol \propto stands for proportionality, and with

$$\mu_{c,t}^2 = \left(\sigma_c^{-2} + \tau^{-2} \{L'(t|x, \theta)\}^{-2}\right)^{-1},$$

$$\eta_{c,t} = \mu_{c,t}^2 \left[\frac{L^*(t)}{\sigma_c^2} + \frac{L(t|x, \theta)}{\tau^2 \{L'(t|x, \theta)\}^2} \right],$$

$$\lambda_{c,t}^2 = \frac{1}{2} \left(\frac{L^*(t)}{\sigma_c} \right)^2 + \frac{1}{2} \left(\frac{L(t|x, \theta)}{\tau L'(t|x, \theta)} \right)^2.$$

The first term in (1.6) corresponds to the distribution of the second observation (recapture) conditioned on the first observation (capture). Denote

$$\begin{aligned} \delta_t^* &= L^*(t + \Delta_t) - L^*(t) && \text{the observed difference,} \\ \delta_t(x, \theta) &= L(t + \Delta_t|x, \theta) - L(t|x, \theta) && \text{the simulated difference} \\ \delta'_t(x, \theta) &= L'(t + \Delta_t|x, \theta) - L'(t|x, \theta). \end{aligned}$$

Since (simplifying the notation $\delta_j = \delta_{t_j}$ and $\delta_j^* = \delta_{t_j}^*$)

$$L^*(t_j + \Delta_j) = L^*(t_j) + \delta_j(x, \theta) + \epsilon_j^c + \epsilon_j^r, \quad (1.7)$$

hence

$$\begin{aligned} [L^*(t_j + \Delta_j)|L^*(t_j), x, \theta, t_j^*, \tau, \sigma^c, \sigma^r] &= \\ \int_0^\infty [L^*(t_j + \Delta_j)|L^*(t_j), \delta_j(x, \theta), \sigma^c, \sigma^r] d[\delta_j(x, \theta)|t_j^*, \tau]. \end{aligned}$$

The first integrand term is provided by the observational equation (1.7).

Based on the same kind of first-order approximation (which is more defensible since at largest sizes length increasing is weak),

$$\delta_j(x, \theta) \simeq \delta_{j^*}(x, \theta) + (t_j^* - t_j) \delta'_{j^*}(x, \theta),$$

it comes (approximately)

$$\left[\delta_j(x, \theta) | t_j^*, \tau'^2 \right] \sim \mathcal{N} \left(\delta_{j^*}(x, \theta), \tau^2 \delta'_{j^*}{}^2(x, \theta) \right).$$

Hence, by conjugation again, it can be found

$$\begin{aligned} [L^*(t_j + \Delta_j) | L^*(t_j), x, \theta, t_j^*, \tau, \sigma_c, \sigma_r] &\propto \\ &\int_0^\infty \frac{1}{\sqrt{\sigma_c^2 + \sigma_r^2}} \exp \left(-\frac{1}{2(\sigma_c^2 + \sigma_r^2)} \{ \delta_j^* - \delta_j(x, \theta) \}^2 \right) \\ &\times \frac{1}{\tau |\delta'_{j^*}(x, \theta)|} \exp \left(-\frac{1}{2\tau^2 \delta'_{j^*}{}^2(x, \theta)} \{ \delta_j(x, \theta) - \delta_{j^*}(x, \theta) \}^2 \right), \\ &\propto \frac{\mu_{r, t_j^*}}{\sqrt{\sigma_c^2 + \sigma_r^2} \tau |\delta'_{j^*}(x, \theta)|} \exp \left(\frac{\nu_{r, t_j^*}^2}{2\mu_{r, t_j^*}^2} - \lambda_{r, t_j^*}^2 \right) \end{aligned}$$

with

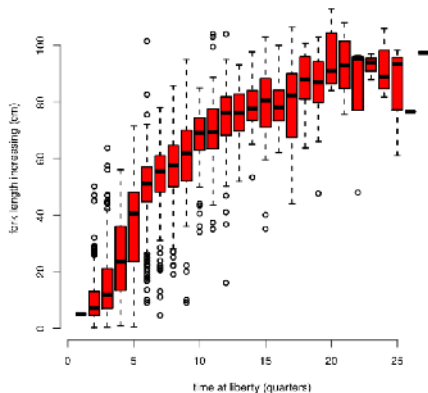
$$\mu_{r, t}^2 = \left(\left[\sigma_c^2 + \sigma_r^2 \right]^{-1} + \tau^{-2} \{ \delta'_t(x, \theta) \}^{-2} \right)^{-1},$$

$$\eta_{r, t} = \mu_{c, t}^2 \left[\frac{\delta_t^*}{\sigma_c^2 + \sigma_r^2} + \frac{\delta_t(x, \theta)}{\tau^2 \delta'_t{}^2(x, \theta)} \right],$$

$$1 - (\delta^*)^2 \quad 1 - (\delta(x, \theta))^2$$

Données dispos (interrogation base RTTP-IO)

5860 yellowfin capture-recapture data



Autre données possibles ?

- ▶ Données larvaires bibliographiques

$$L^*(t_i) = L(t_i|x, \theta) + \xi_i \text{ with } \xi_i \sim \mathcal{N}(0, \sigma^2), \quad (1.8)$$

where $L(t_i|x, \theta)$ is the output of the numerical model. Their corresponding likelihood is

$$\ell_1(x, \theta, \sigma) = \prod_{i=1}^{n_1} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (L^*(t_i) - L(t_i|x, \theta))^2 \right\} \quad (1.9)$$

Un algorithme MCMC classique (random walk)

Step 0:

Sample $X^{(0)} \sim f(X)$ and $\theta^{(0)} \sim \pi(\theta)$

Step i+1: (pour $i \geq 0$)

1. For $k = 1, \dots, d$

(a) sample $\tilde{x}_k^{(i)} \sim \rho_x(x_{k-1}^{(i)}, \lambda_k \{x_{k-1}^{(i)}\}^2)$;

(b) compute the ratio

$$\alpha_{i,k} = \frac{\ell(\mathbf{D}^* | x_1^{(i-1)}, \dots, x_{k-1}^{(i-1)}, \tilde{x}_k^{(i)}, x_{k+1}^{(i-1)}, \dots, x_d^{(i-1)}, \theta^{(i-1)})}{\ell(\mathbf{D}^* | x_1^{(i-1)}, \dots, x_{k-1}^{(i-1)}, x_k^{(i-1)}, x_{k+1}^{(i-1)}, \dots, x_d^{(i-1)}, \theta^{(i-1)})} \frac{f_k(\tilde{x}_k^{(i)}) \rho_x(x_{k-1}^{(i)}, \lambda_k \{x_{k-1}^{(i)}\}^2)}{f_k(x_k^{(i-1)}) \rho_x(x_{k-1}^{(i-1)}, \lambda_k \{x_{k-1}^{(i-1)}\}^2)}$$

(c) select $x_k^{(i)} = \tilde{x}_k^{(i)}$ with probability $\min(1, \alpha_{i,k})$ else fix $x_k^{(i)} = x_k^{(i-1)}$.

2. For $k = 1, \dots, q$

3. sample $\tilde{\theta}_k^{(i)} \sim \rho_\theta(\theta_{k-1}^{(i)}, \lambda'_k \{\theta_{k-1}^{(i)}\}^2)$;

4. compute the ratio

$$\beta_{i,k} = \frac{\ell(\mathbf{D}^* | x^{(i)}, \theta_1^{(i-1)}, \dots, \theta_{k-1}^{(i-1)}, \tilde{\theta}_k^{(i)}, \theta_{k-1}^{(i-1)}, \dots, \theta_d^{(i-1)})}{\ell(\mathbf{D}^* | x^{(i)}, \theta_1^{(i-1)}, \dots, \theta_{k-1}^{(i-1)}, \theta_k^{(i-1)}, \theta_{k-1}^{(i-1)}, \dots, \theta_d^{(i-1)})} \frac{\pi_k(\tilde{\theta}_k^{(i)}) \rho_\theta(\theta_{k-1}^{(i)}, \lambda'_k \{\theta_{k-1}^{(i)}\}^2)}{\pi_k(\theta_k^{(i-1)}) \rho_\theta(\theta_{k-1}^{(i-1)}, \lambda'_k \{\theta_{k-1}^{(i-1)}\}^2)}$$

5. select $\theta_k^{(i)} = \tilde{\theta}_k^{(i)}$ with probability $\min(1, \beta_{i,k})$ else fix $\theta_k^{(i)} = \theta_k^{(i-1)}$.

Un modèle-jouet à 1, 2 ou 3 stances

$$L(t) = \frac{L_d}{1 + \exp(-k_0(t - a_0))} + (L_\infty - L_d) \left\{ 1 - \exp(-k_2(t - a_0)) \left[\frac{1 + \exp(-\beta(t - a_0 - \alpha))}{1 + \exp(\beta\alpha)} \right] \right\}.$$

where

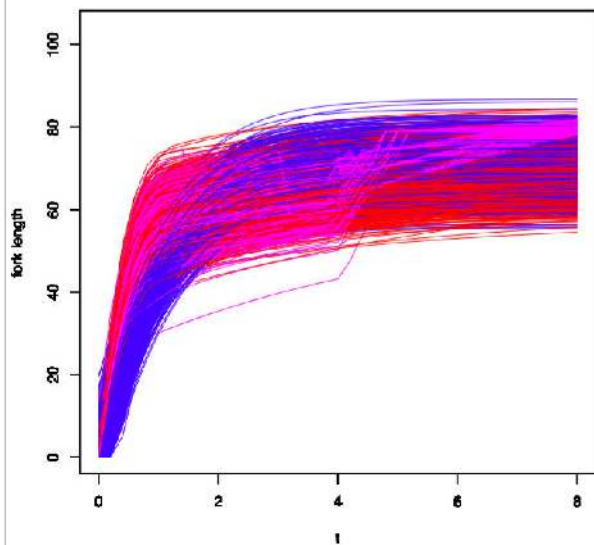
- ▶ L_∞ is the asymptotic length
- ▶ α is the inflection point, i.e the relative age to a_0 at which the change in growth occurs
- ▶ β is the parameter that controls the rate of transition between k_1 and k_2
- ▶ L_d is a transition length, appearing at age $t_d > t_0$, before which the growth is guided by a rate $k_0 < k_2$

If $k_2 \ll k_1$ and $L_d = 0$, the growth curve presents two clear stanza. It reduces to the usual Von Bertalanffy (one-stanza) curve if $k_1 = k_2$

| | L_∞ (cm) | a_0 (y) | α | β | k_1 | k_2 | k_0 | L_d (cm) | t_d (y) |
|-----|-----------------|-----------|----------|---------|-------|-------|-------|------------|-----------|
| | 70.5 | 0 | 0.8 | 18.9 | 2 | 0.35 | 0.2 | 20 | 4 |
| Std | 4.9 | 0.1 | 0.15 | 4.14 | 0.2 | 0.08 | 0.05 | 4.5 | 0.3 |

Table : Input distributions for the toy model (in the encompassing three-stanza case), coming from typical values found for the Indian Ocean skipjack tuna (Eveson et al. 2015).

Essai (2/3)



Essai (3/3) : essai de classification par ACP

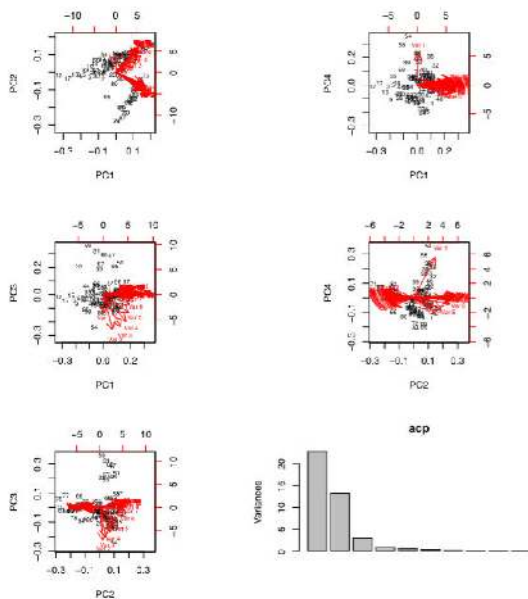


FIGURE 4 Results of a vectorial principal components analysis (PCA) trying to highlight the differences between 15,000 growth curves (5,000 are simulated for each sampling model.)

Problématiques : tour d'horizon (1/2)

- ▶ Recalibration de l'exemple-jouet en fonction des dernières connaissances sur le yellowfin
- ▶ Vérification vraisemblance approximée (en cours avec Agnès Lagnoux)
- ▶ Recalibration fine de l'algorithme d'inversion ; si jamais ça ne fonctionne pas, améliorer le calcul de vraisemblance dans l'algo MCMC (sans doute via une approche par importance sampling)
- ▶ Réduction du temps de calcul via Gibbs par bloc (+ métamodèle code stochastique ?)
- ▶ Simulation du modèle DEB à vérifier
- ▶ Données de capture-recapture réelles à vérifier + reconstitution des âges de première capture
- ▶ Ajout de données larvaires ?
- ▶ Sélection des paramètres du DEB : différencier paramètres environnementaux X et intrinsèques θ
- ▶ Choix des lois a priori pour les paramètres du DEB
- ▶ Analyse de sensibilité a priori du DEB
- ▶ Valeurs des bruits de mesure et bruit de reconstitution sur l'âge (otolithes)
- ▶ Calibration par inversion du DEB sur les données réelles

Problématiques : tour d'horizon (1/2)

- ▶ Analyse de sensibilité a posteriori et classification des situations à 1, 2 ou 3 stances selon la variation des paramètres environnementaux X
- ▶ Calage d'un métamodèle pratique du DEB en fonction des contraintes existantes
 - ▶ Contraintes de forme (monotonie par rapport à certains paramètres d'entrée..) ?
 - ▶ La courbe de longueur est assimilable, pour peu qu'on estime la longueur asymptotique, à une fonction de répartition. Peut-on envisager un métamodèle fondé sur un mélange de fonctions de répartitions ?

