

Poisson-Gamma versus Delta-Gamma pour traiter des données de biomasse à forte proportions de zéros avec un effort d'échantillonnage variable.



Jean-Baptiste Lecomte¹, Liliane Bel¹, Marie-Pierre
Etienne¹, Sophie Ancelet² Hugues Benoît³, Eric Parent¹

Juin 2013

1. UMR AgroParisTech/INRA 518, Équipe MORSE, Paris, France

2. Institut de Radioprotection et de Sécurité Nucléaire, Laboratoire
d'épidémiologie, Fontenay-aux-Roses, France

3. Gulf Fisheries, Moncton, New Brunswick, Fisheries and Oceans Canada



Contexte et Objectifs

- ▶ Analyse de données à forte proportions de zéros
- ▶ Modèles statistiques spécifiques



Comment prendre en compte un effort d'échantillonnage variable ?



Modèle Delta-Gamma

Présence : régression logistique

$$X_s \sim \text{Bernoulli}(p_s)$$

$$\text{logit}(p_s) = \alpha_1 + \beta_1(\text{Effort}_s)$$

- ▶ X : présence ou absence
- ▶ p : probabilité de présence

Biomasse positive : si $X_s = 1$

$$Y_s^{\text{pos}} \sim \text{Gamma}(a, b)$$

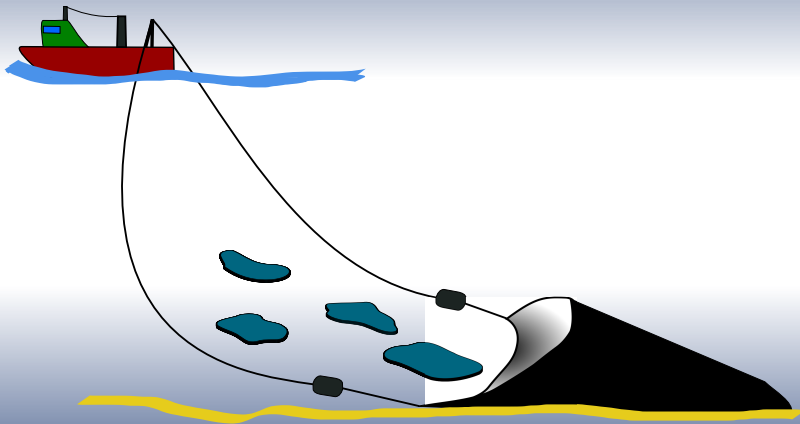
$$E(Y_s^{\text{pos}}) = \frac{a_s}{b} = \alpha_2 + \beta_2(\text{Effort}_s)$$

- ▶ Y_s^{pos} : Biomasse positive
- ▶ $\frac{a_s}{b}$: Biomasse positive moyenne

Biomasse moyenne attendue : $E(Y) = p_s * \frac{a_s}{b}$



Un modèle alternatif : Poisson Composé Gamma





Poisson Composé Gamma

Processus latent

Nombre de patches : $N_s \sim \text{Poisson}(\mu_s)$

Quantité de biomasse : $M_{s,i} \sim \text{Gamma}(a, b)$



Poisson Composé Gamma

Processus latent

Nombre de patches : $N_s \sim \text{Poisson}(\mu_s)$

Quantité de biomasse : $M_{s,i} \sim \text{Gamma}(a, b)$

Modèle des observations

$$Y_s = \begin{cases} \sum_{i=1}^{N_s} M_{s,i} & \text{si } N_s > 0 \\ 0 & \text{si } N_s = 0 \end{cases}$$



Poisson Composé Gamma

Processus latent

Nombre de patches : $N_s \sim \text{Poisson}(\text{Effort}_s \mu_s)$

Quantité de biomasse : $M_{s,i} \sim \text{Exp}(\rho)$

Modèle des observations

$$Y_s = \begin{cases} \sum_{i=1}^{N_s} M_{s,i} & \text{si } N_s > 0 \\ 0 & \text{si } N_s = 0 \end{cases}$$



Comment construire un jeu de données sans utiliser les deux modèles proposés

- 1 Simulation d'un grand nombre d'éléments unitaires (ou micro-sites) dans une loi Delta-Gamma (α_{micro} , β_{micro} et π_{micro})
 - 2 Somme d'éléments unitaires pour obtenir une valeur simulée
- Nombre d'éléments à sommer pour obtenir une "observation simulée" :

$$N_{micro} = \frac{Effort_{obs}}{Effort_{micro}} \quad \text{avec} \quad Effort_{micro} = 0.001$$



Exemple d'un jeu de données simulé

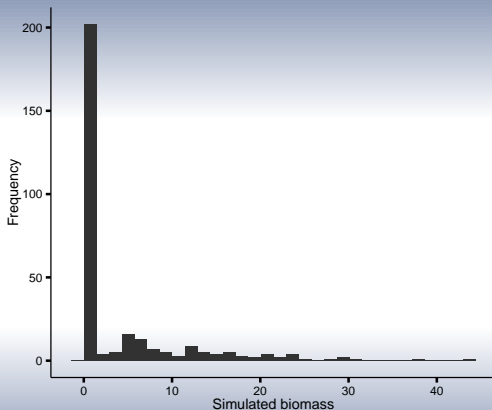
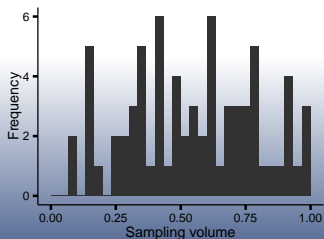
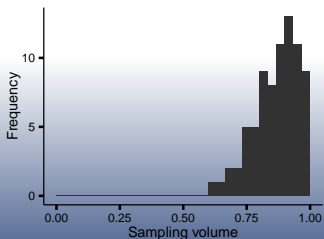
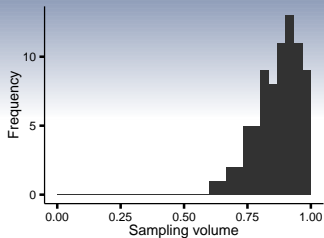
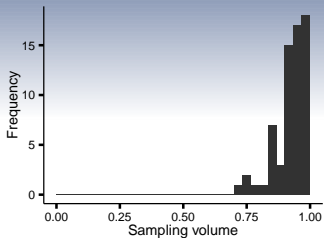


FIGURE : Histogramme de biomasse simulée avec $Effort_{obs} = 1$



Simulation d'un jeu de données à forte proportion de zéros

Effort d'échantillonnage variable



Modélisation de données à forte proportion de zéros avec échantillonnage variable

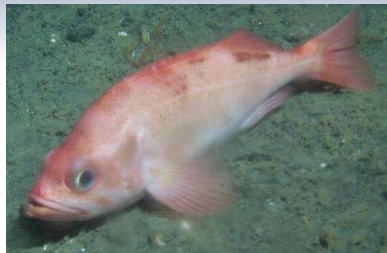


Données commerciales poissons de fonds



Microstomus pacificus

$$2008 \hat{\sigma}_{\text{Effort}}^2 = 0.20 \text{heure}^2$$



Sebastes alutus

$$2009 \hat{\sigma}_{\text{Effort}}^2 = 0.14 \text{heure}^2$$

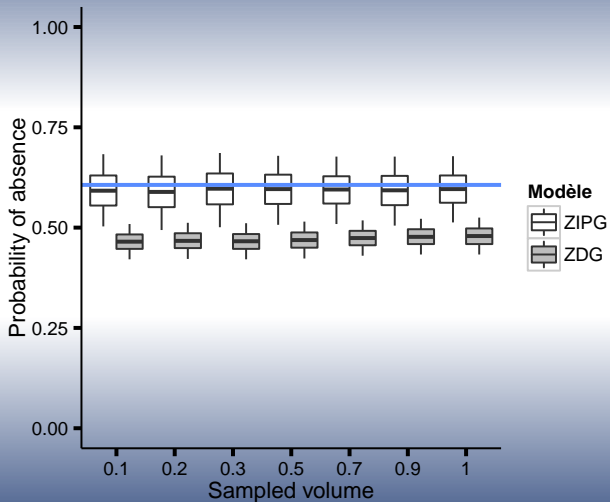


Inférence Bayésienne

- ▶ Méthodes MCMC
- ▶ OpenBUGS
- ▶ Diagnostic de convergence
(30,000 burn-in + 50,000 itérations)

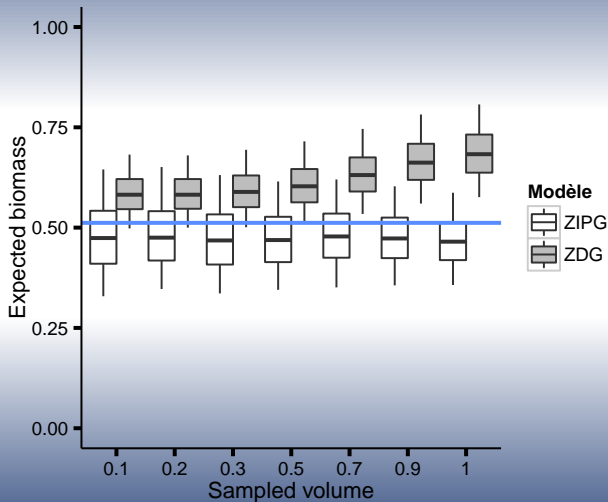


Probabilité d'absence



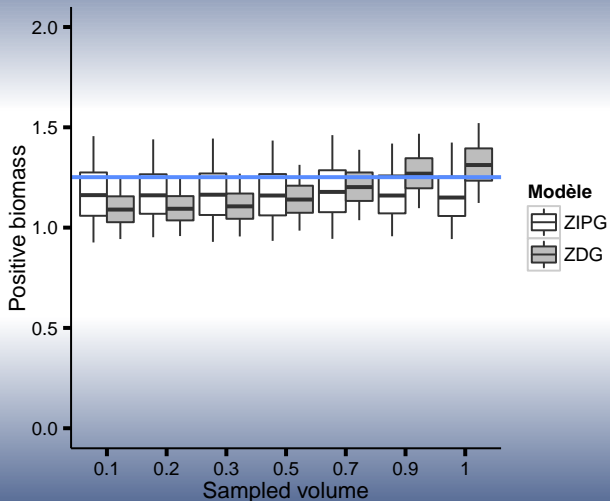


Quantité de biomasse





Quantité de biomasse positive

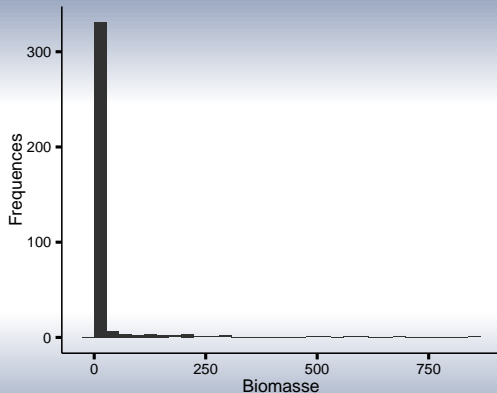




Données commerciales poissons de fonds



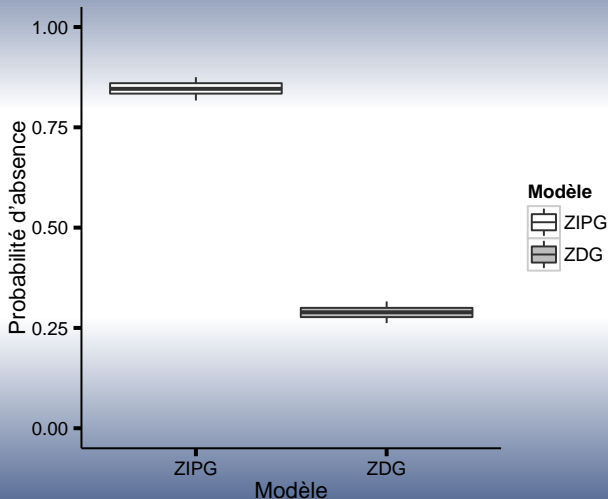
*Microstomus
pacificus*



$$2008 \hat{\sigma}_{Effort}^2 = 0.20 \text{heure}^2$$

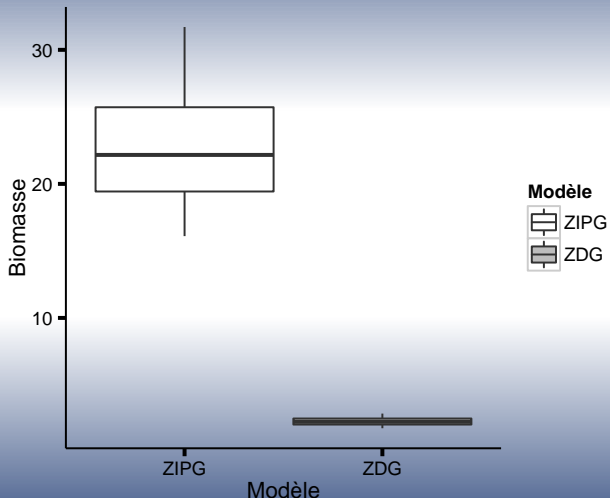


Probabilité d'absence



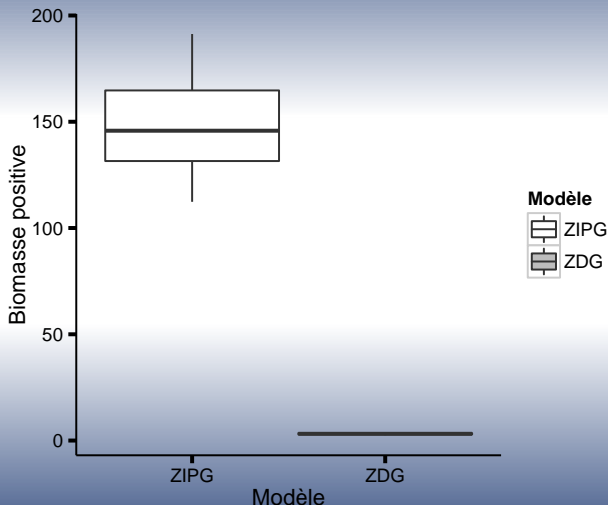


Quantité de biomasse





Quantité de biomasse positive





Recommandations

- ▶ Effort d'échantillonnage constant ou peu variable → Delta-Gamma
- ▶ Effort d'échantillonnage variable → Poisson Composé Gamma

Merci de votre attention