



HBM for Ecological Data

Reprise sous STAN, intérêt et limites

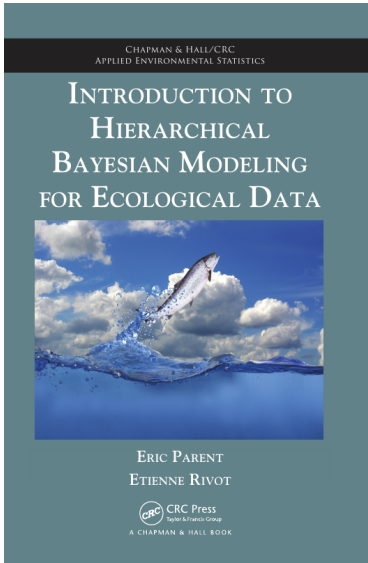
Saint-Clair Chabert-Liddell

23 septembre 2022

Séminaire scientifique E. Parent, IHP

UMR MIA Paris-Saclay Université Paris-Saclay, INRAE, AgroParisTech

HBM for Ecological Data



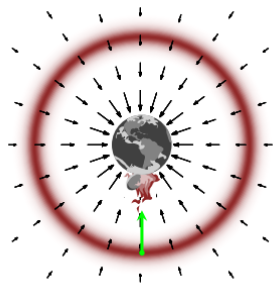
5 types de modèles

- Capture-Marquage-Recapture (CMR)
- Enlèvements successifs (SucRem)
- Stock-Recrutement (StockRec)
- Production de biomasse (Biomass)
- Cycle de vie (Cycle)

Site internet existant <http://www.hbm-for-ecology.org>

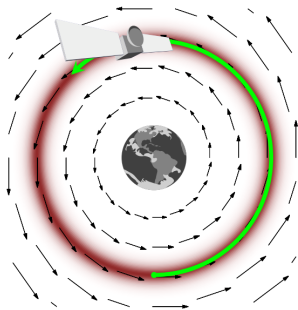
- Code disponible en Bugs
- Sortie graphique en R base

Objectif : Reprendre les modèles en Stan, moderniser les sorties graphiques



En grande dimension

- Visiter la masse autour du mode
- Suivre le gradient mène au mode
- Introduire le bon momentum



- Pour chaque θ_d , variable auxiliaire r_d
- $p(\theta, r) \propto \exp(\ell(\theta) - \frac{1}{2}r \cdot r)$
- Stan utilise l'algorithme No U-Turn Sampling (Hoffman et al., 2014, NUTS)

On différencie par rapport à θ

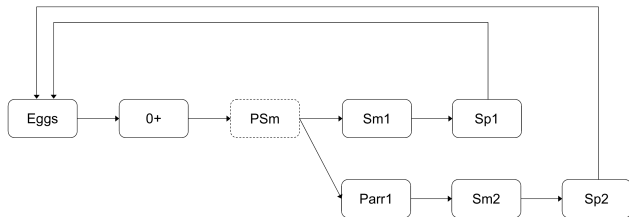
-
- . Betancourt (2017)
 - . Carpenter et al. (2017)

Paramètres discrets

Différenciation impossible avec des paramètres discrets (par ex. : nombre de tirages d'une loi Binomiale)

- Reparamétrisation : possible sur modèle de Capture-Recapture simple
- Marginalisation : difficile pour CMR, SucRem, Cycle
- Approximations normales : dépend des données
- Rendre certaines transitions déterministes

Le cycle de vie du saumon



- œufs → juvéniles → pré-smolts après un hiver
- Les pré-smolts migrent en mer ou passent un an de plus en rivière.
- Les smolts mûrissent en spawners après 1 ou 2 hivers en mer.
- Les spawners retournent en rivière âgés de 2 à 4 ans.

Observations

- Capture-marquage-recapture de *smolts* avant leur départ en mer

$$C_t^{Sm} \sim Bin(Sm_t, \pi_t^{Sm})$$

$$R_t^{Sm} \sim Bin(M_t^{Sm}, \pi_t^{Sm})$$

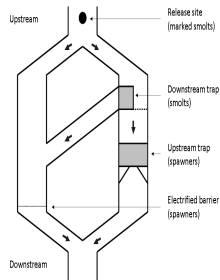
- Capture-marquage-recapture de *spawners* lors de leur retour en rivière

$$C_t^{Sp} \sim Bin(Sp_t, \pi_{1,t}^{Sp})$$

$$Rm_t^{Sp} \sim Bin(M_{2,t}^{Sp}, \pi_{2,t}^{Sp})$$

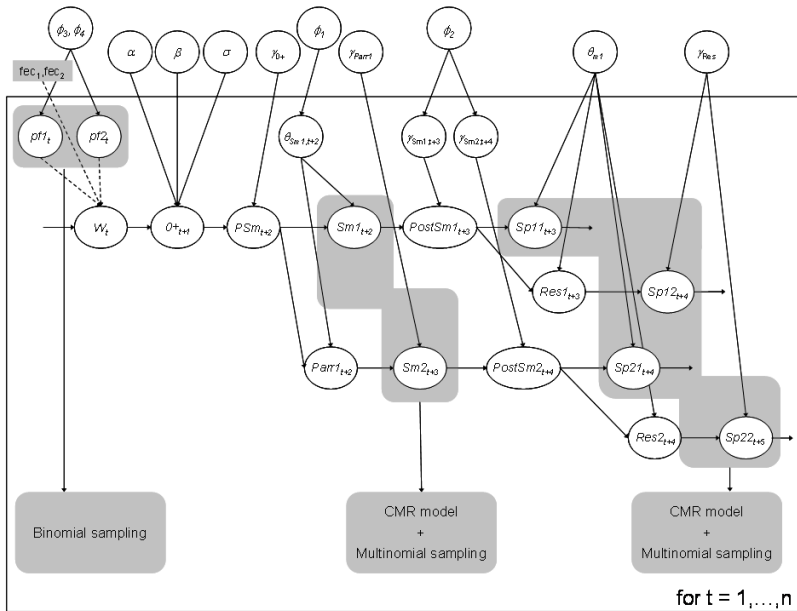
$$Ru_t^{Sp} \sim Bin(Sp_t - C_t^{Sm}, \pi_{2,t}^{Sp})$$

- Informations sur l'âge et le sexe





Le cycle de vie du saumon



- Transition d'état Binomiale

Solution Transition d'état déterministe

Conséquence La variabilité se déplace dans les paramètres hiérarchiques

- Modèle CMR avec paramètres discrets

Solution Approximation normale

Problème Impossible suivant les données

- Capture binomiale à partir d'observations pour renforcer une probabilité

Solution Aucune adaptation nécessaire

Problème Peut forcer certains paramètres à être proche des bords

- Captures très variable suivant les années
- Approximation normale pas raisonnable pour certaines années
- Certaines années aucun *spawner* non marqué n'est capturé ($Ru_t^{Sp} = 0$)
 - Force certains paramètres de binomiale à être proche de 1 ($\pi_{1,t}^{Sp}$)
 - Stan reste bloquer trop près des bords de l'espace des paramètres et n'arrive plus à parcourir l'espace

- Captures très variable suivant les années
- Approximation normale pas raisonnable pour certaines années
- Certaines années aucun *spawner* non marqué n'est capturé ($Ru_t^{Sp} = 0$)
 - Force certains paramètres de binomiale à être proche de 1 ($\pi_{1,t}^{Sp}$)
 - Stan reste bloquer trop près des bords de l'espace des paramètres et n'arrive plus à parcourir l'espace

$$Ru_t^{Sp} \sim \text{Bin}(Sp_t - C_t^{Sm}, \pi_{2,t}^{Sp})$$

$$Rm_t^{Sp} \sim \text{Bin}(M_{2,t}^{Sp}, \pi_{2,t}^{Sp})$$

$$C_t^{Sp} \sim \text{Bin}(Sp_t, \pi_{1,t}^{Sp})$$

- Captures très variable suivant les années
- Approximation normale pas raisonnable pour certaines années
- Certaines années aucun *spawner* non marqué n'est capturé ($Ru_t^{Sp} = 0$)
 - Force certains paramètres de binomiale à être proche de 1 ($\pi_{1,t}^{Sp}$)
 - Stan reste bloquer trop près des bords de l'espace des paramètres et n'arrive plus à parcourir l'espace

Conclusion

- Exemple du livre codé en Jags
- Stan échantillonne très bien à partir de données trafiquées

Modèles

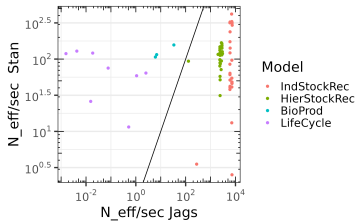
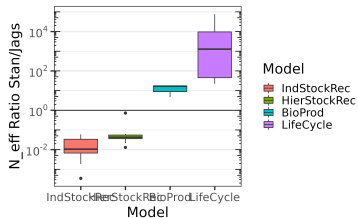
- + Aucune adaptation nécessaire : stock-recrutement, production de biomasse
- ± Dépend des données : CMR, cycle de vie
- Sans trop d'espoir : enlèvement successif

Stan en pratique

- + Super diagnostics de convergences
- + Robuste à l'initialisation (sur ces modèles)
- Syntaxe `rstan` et `cmdstanr` différentes (v2.22 Vs v.2.30)

Comparaison Jags/Stan

4 modèles : Stock-Recrutement (hiérarchique et indépendant),
Production de biomasse, Cycle de vie (données modifiées)



- Performances de Jags s'effondrent avec la complexité du modèle
- Stan plus lent pour les modèles simples mais performances stables



<http://www.hbm-for-ecology.org/>



Merci Éric !

-
- Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv :1701.02434*.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan : A probabilistic programming language. *Journal of statistical software*, 76(1).
- Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler : adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1) :1593–1623.
- Parent, E., Rivot, E., and Rivot, E. (2013). *Introduction to hierarchical Bayesian modeling for ecological data*. CRC press Boca Raton, Florida.