Modèle poissonnien avec sur-représentation de zéros et dépendance spatiale : application à l'étude de la répartition spatiale de juveniles en forêts tropicales humides

Frédéric Mortier & Olivier Flores

CIRAD; unité génétique Forestière

Plan

- Contexte.
- Sur-représentation de zeros et modèle de mélange : *Zero Inflated Model*.
- Prise en compte de la dépendance spatiale.
- Estimation et implémentation WinBUGS.
- Application.
- Conclusions et perspectives.

Contexte

Exploitation durable des forêts

- ☞ Quelle vitesse de reconstitution des stocks ?
- ☞ Processus de régénération ?
  - ➾ processus biologiques impliqués ?
  - ➾ quel impact de l'environnement ?

Richesse spécifique Forêts tropicales

☞ 50% des espèces ; $+$ de 100 espèces à l'hectare

➠ compréhension difficile des écosystèmes : construction de groupe d'espèces fondés sur les stades adultes.

- besoin en lumière
- stratégie de dispersion

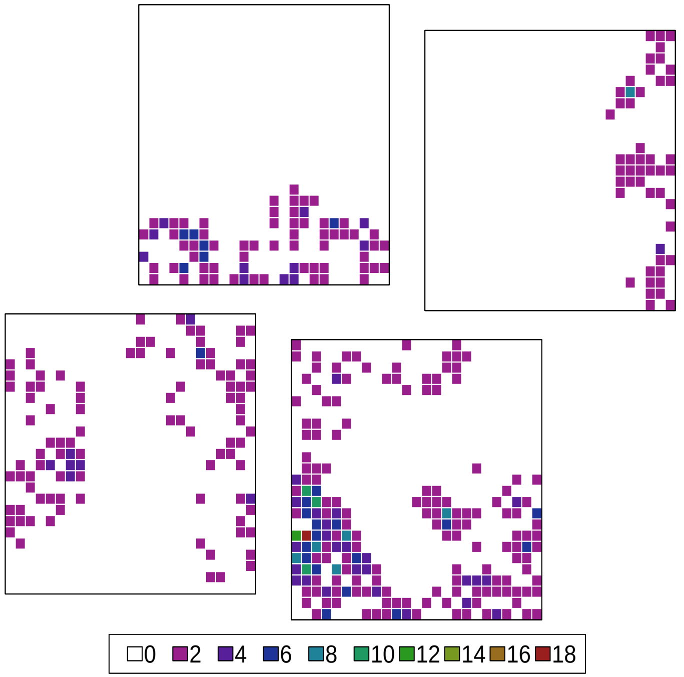
☞ stades juvéniles peu ou mal connus :

➠ Quelles sont les conditions environnementales qui expliquent la répartition spatiale des stades juvéniles ?

➠ les groupes fondés sur les adultes sont-ils valables pour les stades jeunes ?

Site expérimental et observations

- ☞ Paracou en Guyane française
- ☞ nombre d'individus d'une espèce dont la taille $\in [1, 10cm]$ de dbh sur des quadrats 10×10



- ☞ conditions environnementales pour chaque quadrat résumées en 15 variables environnementales.

☞ Approche classique pour des données de comptage : modèle de Poisson.

☞ Mais le nombre de quadrats vides sur représenté : sur-dispersion :
mauvais ajustement d'un modèle de Poisson.

une proportion p des observations est issue

d'une poisson d'espérance nulle (masse de dirac en zero) et

une proportion $(1-p)$ issue d'une poisson d'intensité μ

Zero Inflated Model

Z suit un *Zeros Inflated Modèle*, $Z \sim ZIP(p, \mu)$

si

$$\mathbb{P}(Z = z|p, \mu) = \begin{cases} p + (1-p)\mathbb{P}(Z = 0|\mu), & \text{si} \quad z = 0 \\ (1-p)\mathbb{P}(Z \neq 0|\mu), & \text{si} \quad z > 0 \end{cases}$$

$$\Updownarrow$$

$$f(z|p, \mu) = pP(0) + (1-p)P(\mu)$$

$$\mathbb{E} = (1-p)\mathbb{E}_P(Z|\mu) \qquad \mathbb{V} = p(1-p)\left[\mathbb{E}_P(Z|\mu)\right]^2 + (1-p)\mathbb{V}_P\left(Z|\mu\right)$$

Vraisemblance

Soit $C = (C_1, \ldots, C_n)$ un vecteur latent

C_i prend la valeur $c_i = 0$ si $Z_i > 0$ ou si Z_i nul et issu de $P(\mu)$ et $c_i = 1$ si Z_i nul et issu de $P(0)$.

Marginales $C_i \sim Bern(p)$ $(p = P(C_i = 1))$. la vraisemblance jointe s'écrit alors

$$\begin{aligned} \ell(Z, C|p, \mu) &= \prod_{i=1}^{n} \ell(Z_i|C_i = c_i, p, \mu)\pi(C_i|p) \\ &= \prod_{i=1}^{n} p^{c_i} \left[(1-p)\, P(\mu)\right]^{1-c_i} \end{aligned}$$

Prise en compte de variables explicatives

La présence/absence et l'abondance dépendent de l'environnement :

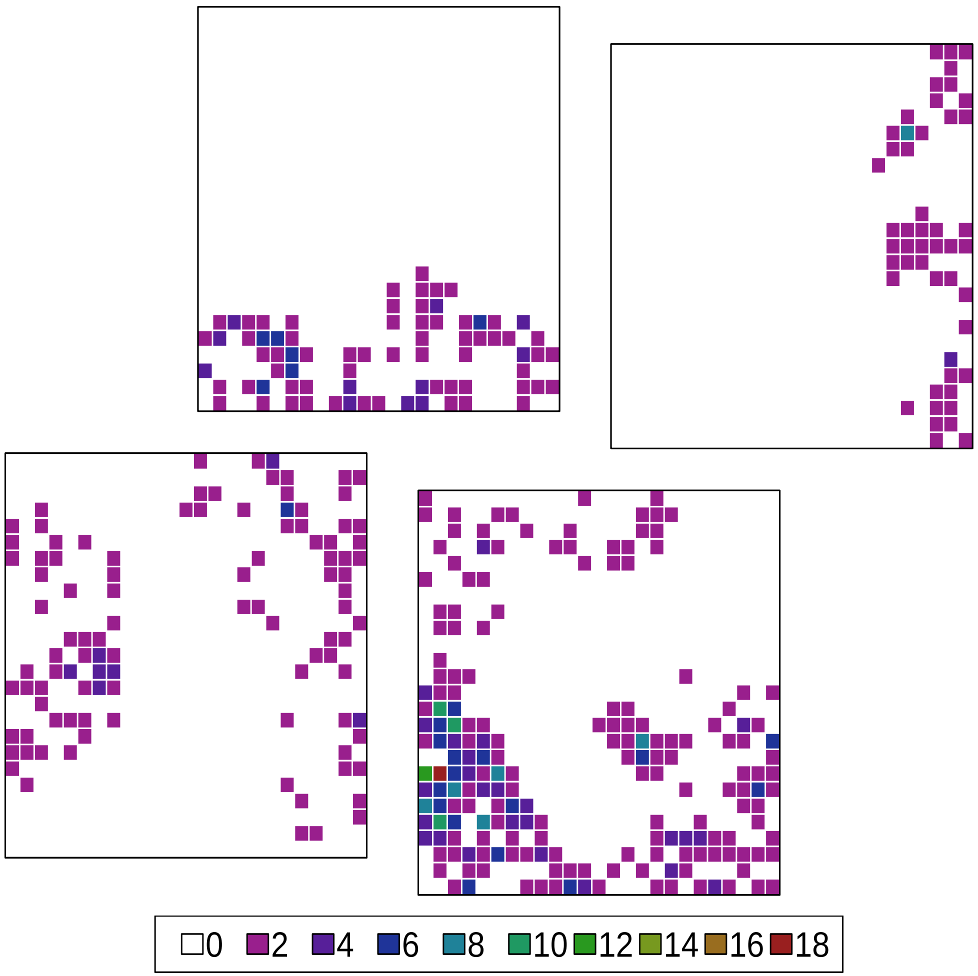
☞ la proportion du mélange dépend de variables environnementales :

$$logit(p_i) = B_i \gamma$$

☞ l'intensité dépend également de variables environnementales :

$$\log(\mu_i) = X_i \beta$$

Prise en compte de la dépendance spatiale



Comment décrire le fait $\mathbb{P}(Z_i, Z_j) \neq \mathbb{P}(Z_i)\mathbb{P}(Z_j)$.

Modèle hierarchique :

1. Data level : $Z_i|p_i, \mu_i \sim ZIP(p_i, \mu_i)$ et $(Z_i|p_i, \mu_i) \perp (Z_j|p_j, \mu_j)$
2. Process Level :
   - proportion du mélange : $(logit(p_i)|\gamma) = B_i\gamma$
   - intensité : $(log(\mu_i)|\beta_i, \alpha_i) = X_i\beta + \alpha_i$
3. Prior : $\beta, \gamma \sim \mathcal{N}(0, 1000)$ et l'effet Spatial : $\alpha_i \sim F(\theta)$
4. hyper-prior $\theta \sim g$:

Composante spatiale

Données à support discret (lattice).

Processus conditional autoregressive (CAR) :

$$\alpha_i | \alpha_j, j \in V_i \sim \mathcal{N}\left(\sum_{j \in V_i} \rho c_{ij} \alpha_j, \sigma^2\right)$$

– V_i voisinage de i (8 voisins)
– $E(\alpha_i) = 0$
– σ^2 : variance conditionnelle
– ρ : paramètre qui pondère la dépendance spatiale
– $C = (c_{ij})$: matrice de poids

$\theta = (\rho, \sigma^2)$

hyper-prior : $\rho \sim U]a, b[, \sigma^2 \sim IG(0.0001, 0.0001)$

Estimation et implémentation Winbugs

simuler loi *a posteriori*

$$
\begin{aligned}
\pi(\beta, C, \gamma, \alpha, \rho, \sigma^2 | Z) \quad &= \quad f(Z|C, \gamma, \beta, \rho, \sigma^2, \alpha)\pi(C|\gamma)\pi(\alpha|, \rho, \sigma^2) \\
&\quad\quad \pi(\beta)\pi(\gamma)\pi(\rho)\pi(\sigma^2).
\end{aligned}
$$

Utilisation de WinBUGS.

Winbugs

- ☞ logiciel libre
- ☞ compatible R (R2WinBUGS ou OpenBUGS)
- ☞ flexible et bien adapté à la modélisation hiérarchique

```
#y = ax + b + e\\
 model{
 for(i in 1 : N){
        y[i]   ~ dnorm(mu{i], tau)
        mu[i] <- a*x[i] + b
    }
 a ~ dnorm(0,0.00001)
 b ~ dnorm(0,0.00001)
 tau ~dgamma(0.0001,0.0001)
 sigma <- 1/tau
 }
```

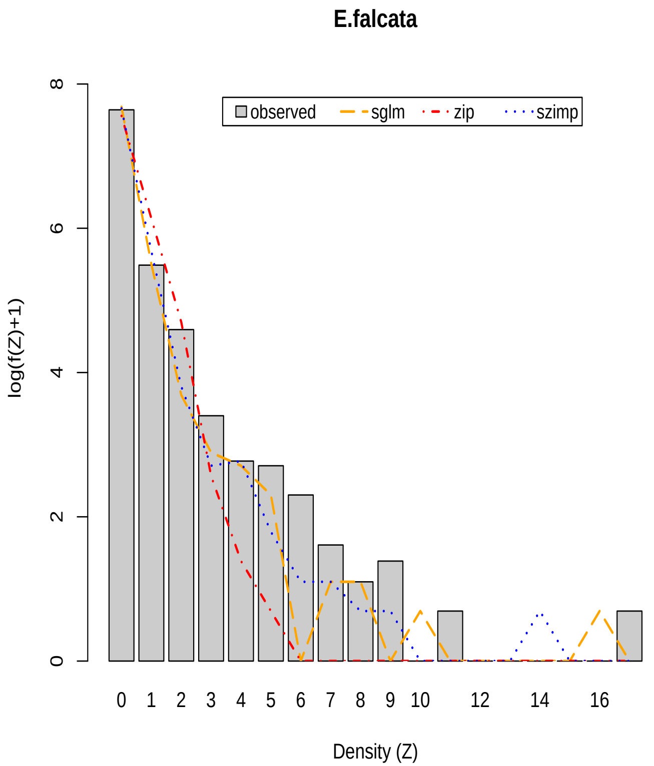
Cas du *ZIP*(ω, μ)

```
model{
        for (i in 1:N){
        Z[i] ~ dpois(lambda[i])
        lambda[i] <- (1 - C[i])*mu[i]
        C[i] ~ dbern(omega[i])
        logit(omega[i]) <- ab[1] + ab[2]*sDL[i]
        log(mu[i]) <- ap[2]*Ax2[i] + ap[3]*Sb[i] + alpha[i]
        }
```

```
    # prior spatial : CAR
    alpha[1:N] ~ car.proper(delta[],C[],adj[],
                  num[],M[],tau.s,gamma)
    # priors LAMBDA
    for(i in  1:2){ap[i] ~dnorm(0, 0.01)}
    # piors OMEGA
    for(i in  1:3){ab[i] ~dnorm(0, 0.01)}
    for(i in 1:N) {delta[i] <- ap[1] }
    #hyper-prior
    tau.s~dgamma(0.1, 0.01)
    var.s<-1/tau.s
    gamma.min <- min.bound(C[],adj[],num[],M[])
    gamma.max <- max.bound(C[],adj[],num[],M[])
    gamma ~ dunif(gamma.min,gamma.max)
}
```

Application

Eperua Falcata ; dispersion non assistée et limitée, hemi-tolérante, n=2500, [0,18] individus par quadrat, 83% de zéros dans les observations. $72,4\%$ attendus $P(0.323)$;

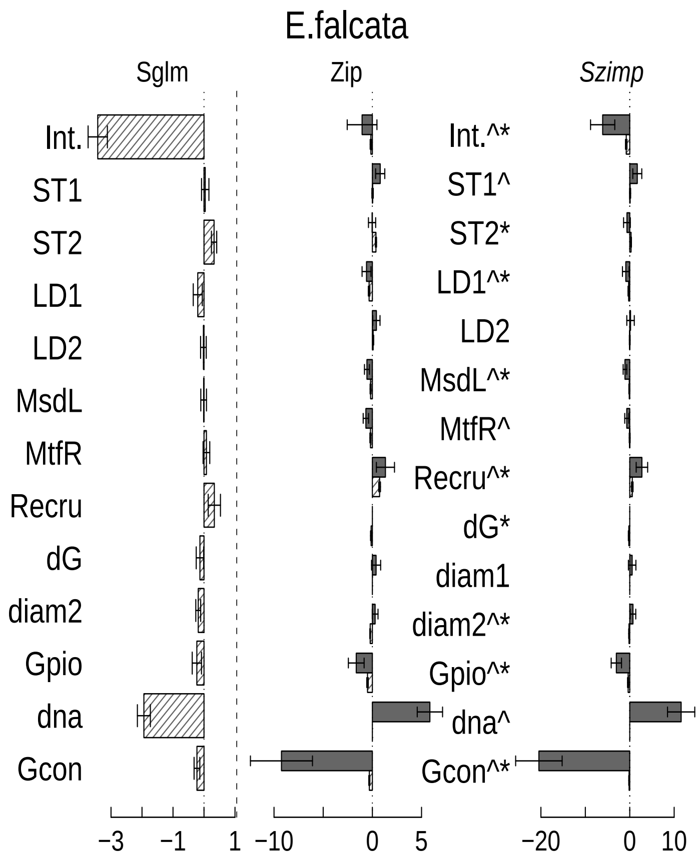


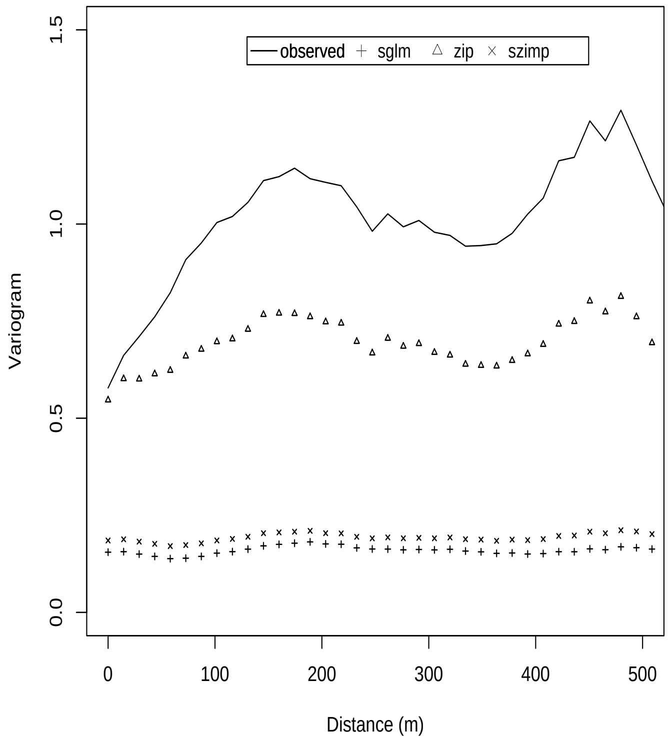
$$DIC$$

$$ZIP = 3067$$

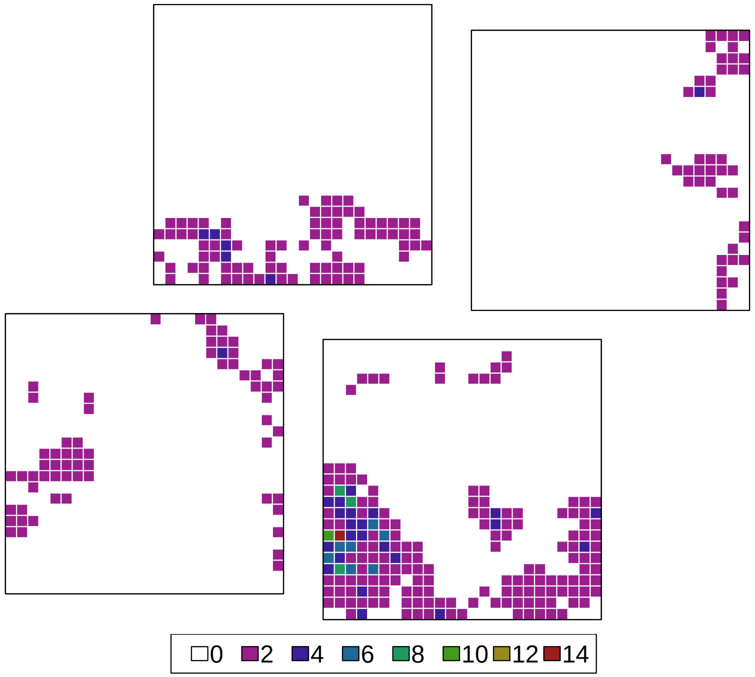
$$Sglm = 2742$$

$$\mathbf{SZIP = 2499}$$


E.falcata
Sglm
Zip
Szimp
Int.
ST1
ST2
LD1
LD2
MsdL
MtfR
Recru
dG
diam2
Gpio
dna
Gcon
Int.^*
ST1^
ST2*
LD1^*
LD2
MsdL^*
MtfR^
Recru^*
dG*
diam1
diam2^*
Gpio^*
dna^
Gcon^*
−3 −1 1
−10 0 5
−20 0 10


observed
sglm
zip
szimp
Variogram
0.0
0.5
1.0
1.5
0
100
200
300
400
500
Distance (m)

0 2 4 6 8 10 12 14 16 18



0 2 4 6 8 10 12 14

Conclusions et perspectives

Conclusions

☞ ZIP tient bien compte de la sur-dispersion

☞ CAR comme *a priori* rend bien compte de la structure spatiale

☞ Approche hiérarchique bayésienne intéressante en ecologie

☞ effet spatial très fort ! pouvoir prédictif en écologie ?

☞ selection variable, RJMCMC ?

Perspectives

- ☞ l’environnement est spatialisé
  - ➩ travailler conditionnellement à l’environnement,
    - ➠ Krigeage environnement
    - ➠ prise en compte de la répartition spatiale des parents
    - ➠ la proportion du mélange devient spatialisé.
- ☞ Modèles au niveau individuel (avec Dessard H. & Picard N.)
- ☞ prise en compte de la structure dynamique des écosystèmes ?