

# SPATIAL BAYESIAN MODELS OF TREE DENSITY WITH ZERO INFLATION AND AUTOCORRELATION

Frédéric MORTIER\*, Olivier FLORES  
and Sylvie GOURLET-FLEURY\*\*

## ABSTRACT

Understanding the spatial and temporal dynamics of rain forests is a challenge for assessing the impact of disturbance on forest stands and tree populations. Still few studies address the modelling of spatial patterns of tree density. Here, we present Hierarchical Bayesian (HB) models for the local density of juveniles trees in a tropical forest. These models are specifically designed to handle zero inflation and spatial autocorrelation in the data.

Height types of models were built and compared through a Hierarchical Bayesian approach: Poisson and Negative Binomial generalized linear models, zero-inflated versions of these models and finally a spatial generalization of the four previous models. Spatial dependency in juvenile pattern was modeled through a Conditional Auto Regressive process.

An application is presented at the Paracou experimental site (French Guiana). At this site, permanent sample plots settled in a previously undisturbed forest received silvicultural treatments in 1986-1988. Juvenile density of a timber species, *Eperua falcata* (Caesalpiniaceae), was evaluated in 2003 within 10 m × 10 m cells and served as response in the models. Explanatory variables described three aspects of environmental heterogeneity inside the plots: topography (elevation and slope) was derived from a Digital Elevation Model; stand variables and population variables, either static or dynamic, were calculated from basal area on 20 m-radius circular subplots.

*Keywords* : Spatial pattern, Hierarchical models, Zero-inflation, MCMC, conditional autoregressive process.

## RÉSUMÉ

Comprendre la dynamique spatio-temporelle des forêts tropicales humides est un défi dans l'évaluation de l'impact des perturbations sur les peuplements et populations d'arbres forestiers. Cependant peu d'études concernent la modélisation des répartitions spatiales de la densité d'arbres. Dans cet article, nous présentons des modèles hiérarchiques bayésiens (HB) de densité d'arbres juvéniles en forêt tropicale. Ces modèles sont construits spécifiquement en prenant en compte la sur-représentation de zéros et l'autocorrélation spatiale dans les observations.

---

\* CIRAD – Diversité génétique et amélioration des espèces forestières, Campus international de Baillarguet, TA 10/C, 34398 Montpellier Cedex 5, France.  
frederic.mortier@cirad.fr

\*\* CIRAD – Dynamique des Forêts Naturelles, Campus International de Baillarguet, TA 10/D, 34398 Montpellier Cedex 5, France.

Huit modèles sont construits et comparés selon une approche hiérarchique bayésienne : les modèles linéaires généralisés utilisant les distributions de Poisson et Négative Binomiale, une version de ces modèles prenant en compte l'excès de zéros, et enfin une généralisation de ces quatre modèles avec autocorrélation spatiale. Un processus Conditionnel Auto-Régressif est utilisé pour modéliser la dépendance spatiale au sein de la répartition des juvéniles.

Une application est présentée sur le site expérimental de Paracou en Guyane française. Sur ce site, des parcelles ont été initialement installées en forêt non perturbée puis ont subi des traitements sylvicoles pendant la période 1986-1988. La densité des juvéniles d'une espèce de canopée exploitée (*Eperua falcata*, Caesalpiniaceae) a été évaluée en 2002 par quadrats de 10 m×10 m et utilisée comme réponse dans les modèles construits. Les variables explicatives utilisées quantifient trois aspects de l'hétérogénéité environnementale au sein des parcelles : la topographie (altitude et pente) a été évaluée à partir d'un modèle numérique de terrain ; des variables de peuplement et de population, statiques et dynamiques, ont été calculées sur des placettes circulaires de 20 m de rayon.

*Mots-clés* : Répartition spatiale, Modèle hiérarchique, Zero-inflation, MCMC, Processus conditionnel autorégressif.

## 1. Introduction

The certification of forests is a process by which loggers become accredited by a seal of quality in exchange of the respect of silvicultural rules compatible with the sustainability of timber production (OAB-OIBT, 2003). In order to propose relevant rules, it is necessary to correctly predict the recovery of the stock after timber logging, that is the dynamics of tree populations, on the mid and long term. Mid term prediction of stock recovery is relatively easy to achieve because the reaction to disturbance of trees above 10 cm diameter at breast height is well documented (Gourlet-Fleury *et al.*, 2004). But poor knowledge of regeneration processes hinders long term prediction: little is known about which factors explain the presence of juveniles at a given place. In particular, the respective part played by physical factors (topography), resources (light, nutrients, water) and relationships with conspecific trees is rarely stated or quantified. In this study, our objective was to model the relationships between juvenile density and those factors in order to predict the spatial pattern of trees.

We used data from the Paracou experimental site (French Guiana) which is dedicated to the study of the impact of silvicultural treatments (logging, logging + thinning) on stands and tree populations dynamics, through a more than 20 years monitoring of 12 large permanent sample plots. We focused on *Eperua falcata* (Caesalpiniaceae), a canopy tree species logged for timber. We used a survey of the juveniles of *Eperua falcata* conducted in 2002-2003. Juvenile density was estimated in 10 m×10 m quadrats and constituted the observations. Each quadrat was characterized by 7 environmental variables describing topography, stand structure and disturbance. The basal area of conspecific trees ( $\geq 10$  cm DBH) was also taken into account.

When observations are count data, Poisson or Negative Binomial models are classically used (McCullagh and Nelder, 1989). But the spatial patterns of tropical tree species are often clumped (Condit *et al.*, 2000), so that the local density of a given life-stage is likely to show spatial autocorrelation (Legendre, 1993). Autocorrelation challenges the common statistical hypothesis of observations being independent. Moreover, quadrat-sampling may also induce dispersion in the data due to many zero counts (*zero inflation*, McCullagh and Nelder, 1989; Ridout *et al.*, 1998). Zero inflation leads to very poor fits of classical models. Alternatively, mixture models formulation can be used to take zero inflation into account.

In this paper we present and compare four models classically used for count data (Poisson, Negative Binomial, Zero Inflated Poisson and Zero Inflated Negative Binomial models) and their spatial generalization to predict the spatial pattern of juveniles in a tropical rainforest. The statistical models were specified in order to handle zero inflation and autocorrelation through a Hierarchical Bayesian approach (HB, Clark, 2005). This method allows to model complex biological data into a series of simpler conditional models (Wikle, 2003; Clark, 2005). Moreover, the Bayesian paradigm offers attractive advantages through its ability to formally incorporate prior knowledge or opinion into model specification, via prior distributions (Banerjee *et al.*, 2003), and allows to take excess of variability into account Clark (2005). Within this framework, we addressed the two following questions: which variables, among physical factors, stand structure, disturbance and intraspecific relationships, are the main determinants of the spatial patterns of juveniles? Which statistical model appears more relevant to model those spatial patterns?

## 2. Statistical Models

### 2.1. Models for count data with zero inflation and spatial correlation

To simplify notations and to be in accordance with the Hierarchical Bayesian approach described in section 2.1.3, we noted in the sections 2.1.1 and 2.1.2, the parametric probability mass function of a random variable  $Z$  with parameter  $\theta$  as  $\mathbb{P}(Z|\theta)$ .

#### 2.1.1. Count data

Count data are classically modeled using Poisson distributions  $\mathcal{P}$  (McCullagh and Nelder, 1989). Let  $Z$  be a random count variable.  $Z$  is distributed as a Poisson model with intensity  $\mu$  if the mass function can be written as:

$$\mathbb{P}(Z = \mathbf{z}|\mu) = \frac{\mu^{\mathbf{z}} e^{-\mu}}{\mathbf{z}!} \quad (1)$$

Expectation and variance are then equal to  $\mu$ . But when overdispersion occurs (*i.e.* expectation is less than variance), fits of Poisson models can be very poor.

Plackett (1981) proposed to use the Negative Binomial distribution ( $\mathcal{NB}$ ), defined as a continuous mixture of Poisson distributions. This definition allows the Poisson mean  $\mu$  to be gamma distributed. More specifically, the Negative Binomial mass function is defined as:

$$\mathbb{P}(Z = \mathbf{z}|\mu, \tau) = \frac{\Gamma(\mathbf{z} + \tau)}{\mathbf{z}!\Gamma(\tau)} \left(\frac{\tau}{\mu + \tau}\right)^\tau \left(\frac{\mu}{\mu + \tau}\right)^{\mathbf{z}}, \quad (\mu, \tau) > 0 \quad (2)$$

for  $\mathbf{z} = \{0, 1, 2, \dots\}$  and  $\Gamma$  the gamma function. Expectation and variance are given by  $\mu$  and  $\mu + \mu/\tau$ . The Negative Binomial distribution approaches a Poisson distribution when  $\tau$  tends to infinity. In the particular case of zero inflation, fits of Negative Binomial models can still be very poor.

### 2.1.2. Zero inflation

Authors proposed the use of a mixture formulation to take zero inflation into account. They proposed the use of Zero Inflated models such as Zero Inflated Poisson models (ZIP) or Zero Inflated Negative Binomial models (ZINB). These models assign a mass of  $\omega$  to extra zeroes and a mass of  $(1 - \omega)$  to a Poisson or negative binomial distribution, where  $0 < \omega < 1$ . Both models are characterized by the following probability function

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

where  $\mathbb{P}$  depends on the considered model (Poisson, eq. 1, or Negative Binomial, eq. 2). Using the mixture formulation, the mass functions respectively equal

$$\mathbb{P}(Z|\omega, \mu) = \omega \times \delta(0) + (1 - \omega)\mathbb{P}(Z|\mu) \quad (3)$$

for the ZIP model, where  $\delta(0)$  is the Dirac distribution at zero, and

$$\mathbb{P}(Z|\omega, \mu, \tau) = \omega \times \delta(0) + (1 - \omega)\mathbb{P}(Z|\mu, \tau) \quad (4)$$

for the ZINB model.

Mixture models can be expressed through latent class random variables, which has many benefits. First, the expression of the likelihood is simplified. Second the models can be easily generalized to take covariables into account. Third, estimation and inference are also simplified. Let  $C$  be a latent class random variable, so that  $C$  equals  $c = 0$  if  $Z > 0$  or if  $Z$  equals zero but drawn from  $\mathcal{P}(\mu)$  (resp.  $\mathcal{NB}(\mu, \tau)$ ), and  $c = 1$  if  $Z$  equals 0 coming from zero-inflated state. The marginal distribution of  $C$  is then a Bernoulli distribution  $C \sim \mathcal{B}(\omega)$  with parameter  $\omega = \mathbb{P}(C = 1)$ , and the joint distribution of  $(Z, C)$  equals

$$\begin{aligned} \mathbb{P}(Z, C|\omega, \mu) &= \mathbb{P}(Z|C = \mathbf{c}, \omega, \mu)\mathbb{P}(C|\omega) \\ &= \omega^c [(1 - \omega)\mathbb{P}(Z|\theta)]^{1-c} \end{aligned} \quad (5)$$

where  $\mathbb{P}(Z|\theta)$  with  $\theta = \mu$  is the Poisson mass function for the ZIP model (eq. 1) and with  $\theta = (\mu, \tau)$  is the Negative Binomial mass function for the ZINB model

(eq. 2). Moreover, the generalization of ZIP and ZINB using covariables and the classical link functions of Generalized Linear Models (McCullagh and Nelder, 1989) induce the following equations

$$\text{logit}(\omega) = \mathbf{B}\gamma \quad (6)$$

and

$$\log(\mu) = \mathbf{X}\beta \quad (7)$$

where  $\mathbf{B}$  and  $\mathbf{X}$  are two matrices of covariables, either equal or not, and  $\gamma$  and  $\beta$  unknown parameter vectors.

### 2.1.3. Spatial correlation

A major issue in spatial modeling is to correctly describe the covariance structure of the data, *i.e.* possible autocorrelation between observations. In this purpose, the Hierarchical Bayesian (HB) approach helps decomposing a complex problem into a series of simpler conditional levels (Banerjee *et al.*, 2003; Wikle, 2003): at a given hypothesis level, inference conditionally relies on hypotheses made at higher levels. Three basic levels constitute a model: a *data* level specifies the conditional distribution of data  $Z$  given parameters and underlying processes; a *process* level specifies the conditional distributions of the processes given their own parameters; a *parameter* level specifies prior distributions for all parameters (Wikle, 2003). Additional levels can specify prior distributions for hyperparameters. Then, following the HB approach, a spatial generalization of Poisson models can be defined as follows:

$$\begin{aligned} \text{data level :} & & Z(\mathbf{s})|\mu(\mathbf{s}) &\sim \mathcal{P}[\mu(\mathbf{s})] \\ \text{process level :} & & \log[\mu(\mathbf{s})]|\beta_{\mathbf{s}}, \alpha(\mathbf{s}) &= \mathbf{X}_{\mathbf{s}}\beta_{\mathbf{s}} + \alpha(\mathbf{s}) \\ \text{parameter level :} & & & \text{priors for } \beta_{\mathbf{s}} \text{ and } \alpha \\ \text{hyperparameter level :} & & & \text{hyperpriors for parameters of } \alpha \end{aligned}$$

where  $\mathbf{s}$  is a known vector of spatial locations at which explicative covariables are measured and fixed effects are to be estimated,  $\mathbf{X}_{\mathbf{s}}$  is the matrix of covariables,  $\beta_{\mathbf{s}}$  is the vector of regression parameters, and  $\alpha(\mathbf{s})$  is a spatial random effect. The hypothesis of conditional independence of observations given parameter  $\mu(\mathbf{s})$  then replaces the usual hypothesis of complete independence. This model is a special case of generalized linear mixed models (GLMM).

The ZIP and ZINB formulations were also extended to include spatial autocorrelation. At the *data level*, we supposed that the response variable  $Z$  was spatial ( $Z = Z(\mathbf{s})$ ), and ZIP distributed:  $Z(\mathbf{s})|\omega(\mathbf{s}), \mu(\mathbf{s}) \sim \mathcal{ZIP}[\omega(\mathbf{s}), \mu(\mathbf{s})]$ . At the *process level*,  $\omega(\mathbf{s})$  and  $\mu(\mathbf{s})$  were linked with covariables through canonical link functions and we defined  $u(\mathbf{s}) = \text{logit}[\omega(\mathbf{s})]$  and  $v(\mathbf{s}) = \log[\mu(\mathbf{s})]$  (McCullagh and Nelder, 1989). Hence, the intensity of the Poisson process in the ZIP model depended on a set of covariables and on an underlying spatial process  $\alpha(\mathbf{s})$

$$u(\mathbf{s})|\gamma_{\mathbf{s}} = \mathbf{B}_{\mathbf{s}}\gamma_{\mathbf{s}} \quad (8)$$

$$v(\mathbf{s})|\beta_{\mathbf{s}}, \alpha(\mathbf{s}) = \mathbf{X}_{\mathbf{s}}\beta_{\mathbf{s}} + \alpha(\mathbf{s}) \quad (9)$$

As for the Poisson case, the subscript indicates that covariables and fixed effects are measured at locations identified by  $\mathbf{s}$ . However, explicit spatial dependency is included through the random effect  $\alpha(\mathbf{s})$  only (see below). The generalization used was identical for the ZINB model.

At the *parameter level*, we used weak or non informative prior Gaussian distributions for regression parameters  $\gamma_{\mathbf{s}}$  and  $\beta_{\mathbf{s}}$ . Given that the observations were located on a discrete grid (lattice with 625 cells, see section 3), the *prior* distribution for the spatial process was defined as a Conditional Autoregressive Process (CAR, Besag, 1974), which is a special case of Markov process. As a neighborhood, we used a Moore neighborhood around each cell (the chess king's move). The spatial process intensity,  $\alpha(\mathbf{s})$ , then followed a conditional Gaussian distribution given intensities in the neighborhood:

$$\alpha(s_i)|\alpha(s_j), j \in v_i \sim \mathcal{N}\left(\rho \sum_{j \in v_i} w_{ij} \alpha(s_j), 1/\sigma\right) \quad (10)$$

where  $\rho$  and  $\sigma$  are two unknown parameters,  $(w_{ij})$  is a set of known spatial weights and  $v_i$  is the neighborhood of  $s_i$  (Banerjee *et al.*, 2003). The spatial weights  $(w_{ij})$  were set proportional to the inverse of the number of neighbors in  $v_i$  and such that  $\sum_j w_{ij} = 1$ . In this model,  $\rho$  measures the strength of the relationship between  $\alpha(s_i)$  and  $\alpha(s_j)$  in  $v_i$ , while  $\sigma$  is the conditional precision of the process ( $1/\sigma$  is the conditional variance). Finally, at the *hyperparameter level*, the prior distribution of  $\rho$  was uniform on a constrained interval (see Banerjee *et al.* (2003) for details) and  $\sigma$  followed a gamma distribution.

## 2.2. Inference and models comparison criterion

In this section, we describe modeling and inference in a complete Bayesian framework. The Bayesian framework allows to specify prior distributions which represent the best guesses about the parameters before information provided by the data is taken into account. In general, given data and model parameters, the Bayesian model specification requires a likelihood function and prior distributions. The *posterior* density of the parameters given the data is then obtained by Bayes' theorem as proportional to the product of the likelihood and the priors. This approach is advantageous in that it allows to take uncertainty on the estimates of the parameters into account. For the ZIP and ZINB models, the *posterior* distributions are given by

$$\pi(\beta_{\mathbf{s}}, C, \gamma_{\mathbf{s}}, \alpha, \rho, \sigma | Z) = f(Z|C, \gamma_{\mathbf{s}}, \beta_{\mathbf{s}}, \alpha) \pi(C|\gamma_{\mathbf{s}}) \pi(\alpha|\rho, \sigma^2) \pi(\beta_{\mathbf{s}}) \pi(\gamma_{\mathbf{s}}) \pi(\rho) \pi(\sigma^2) \quad (11)$$

where  $f$  is either the Poisson density function for the ZIP model or the Negative Binomial density function for the ZINB model,  $\pi(\beta_{\mathbf{s}})$ ,  $\pi(\gamma_{\mathbf{s}})$ ,  $\pi(\alpha|\rho, \sigma^2)$  are *prior* distributions of the regression parameters and spatial effect, and  $\pi(\rho)$  and  $\pi(\sigma^2)$  are *hyperprior* distributions for the parameters of the spatial random process (CAR model, eq. 10).

Model calibration was performed using OpenBUGS and the BRugs package for R Development Core Team (2004) with 100000 iterations on one Monte Carlo Markov chain (MCMC) including an initial burning step of 20000 iterations.

Comparing models in a HB context is not a simple task. The *effective* number of parameters or degrees of freedom is not always clearly defined and can be very different from the actual number of parameters (Spiegelhalter *et al.*, 2002). Hence, common criteria, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), are suspicious. Spiegelhalter *et al.* (2002) proposed a Deviance Information Criterion (DIC) based on the moments of the deviance to compare hierarchical models:  $DIC = \overline{D(\theta)} + p_D$ , where  $\theta$  is the parameter set of the model,  $\overline{D(\theta)}$  the mean of the Bayesian deviance  $D(\theta)$  for all MCMC samples. The effective number of parameters is defined as  $p_D = \overline{D(\theta)} - D(\bar{\theta})$ , where  $\bar{\theta}$  is the mean of all MCMC samples of  $\theta$ .  $p_D$  is thus approximately proportional to the deviance variance and measures the complexity of the model (Spiegelhalter *et al.*, 2002). In this paper, DIC has been used for convenient computation reasons. Nevertheless, Celeux *et al.* (2006) have shown the sensibility of DIC to model parameterization and in particular in missing data problems with unobserved latent variables.

### 3. Application

#### 3.1. Study site

The study was conducted at the Paracou experimental site (5°18' N, 52°23' W) in French Guiana. The site lies in a *terra firme* rain forest of the coastal plain under an equatorial climate. A dry season occurs from August to mid-November. From March to April, a short drier period interrupts the rainy season (Gourlet-Fleury *et al.*, 2004). The physiography of the site shows smooth slopes incised by minor streams. Part of the site is covered by permanently waterlogged areas. The experimental design of the site consists in three blocks of four 300 m × 300 m permanent sample plots with a 25 m inner buffer zone. In each central 250 m × 250 m square, all trees over 10 cm DBH (diameter at breast height) were identified and georeferenced. Since 1984, girth at breast height, standing deaths, treefalls and newly recruited trees over 10 cm DBH have been monitored annually (Gourlet-Fleury *et al.*, 2004). In each block, stands experienced three treatments during the 1986-1988 period combining selective logging of increasing intensity and additional thinning by poison-girdling. One plot per block was left as control. The present work focused on the four plots of the Southern Block (three treated plots and one undisturbed control plot). We defined two periods in order to describe past disturbance and dynamics: the logging period (1986-1989) and the recovery period (1989-2003).

### 3.2. Focal species

*E. falcata* is a canopy timber species whose largest trees reach more than 90 cm DBH. The species is mid-shade-tolerant (Favrichon, 1994; Collinet, 1997) able to germinate in the understorey and to bear shaded conditions. Its dispersal mode is autochory: the seeds are dispersed by pods exploding at maturity in the trees, with a possible secondary dispersal by water. The spatial pattern of the trees  $\geq 10$  cm DBH is highly aggregated (Collinet, 1997; Flores *et al.*, 2006). Trees are aggregated at various scales: they form little aggregates of about 30 m of diameter, inside large clumps of about 100 m of diameter. Within the aggregates, the trees are randomly distributed. The clumps are located along bottomlands. They can be clearly delimited and no isolated tree stands outside the clumps.

### 3.3. Variables

Juvenile 1-4 cm DBH density was evaluated within 10 m  $\times$  10 m cells and served as the response in the models. Ecological descriptors of the environment (see Table 1) either derived from a Digital Elevation Model of the study site (elevation and slope) or were calculated from the basal area of trees  $\geq 10$  cm DBH on 20 m-radius circular subplots centered on the sampling cells. This design was used to take into account neighborhood effects of surrounding trees on the observed juveniles density. Disturbance was quantified by the loss in basal area due to treefalls, standing deaths during the logging period and basal area of recruits (*tFL*, *sD* and *Re*). Distance to the nearest adult (*dna*) measured juvenile spatial dispersal around adults. Finally, the basal area of conspecific trees over 10 cm DBH in 2002 measured intraspecific competition.

TABLE 1. — Ecological variables derived from a DEM (Digital Elevation Model) of Paracou, or from census data of trees  $\geq 10$  cm DBH (*units* in brackets) on 20-m circular subplots centered on sampling cells. The period indicates calculus years: 1986-1988 (logging) or 1989-2002 (recovery). Statistical variables were calculated in 2002.

| Type       | Label                   | Description  | Period   |
|------------|-------------------------|--|----------|
| Topography | <i>Ele</i>              | Elevation (m)  |          |
|            | <i>Slo</i>              | Slope ( $^{\circ}$ )                                       |          |
| Stand      | <i>tFL</i>              | Basal area lost in treefalls ( $m^2$ )                     | logging  |
|            | <i>sD</i>               | Basal area lost in standing deaths ( $m^2$ )               | recovery |
|            | <i>Re</i>               | Basal area of recruits $\geq 10$ cm DBH ( $m^2$ )          |          |
| Population | <i>dna</i>              | Distance to the nearest adult                              | 2002     |
|            | <i>G<sub>cons</sub></i> | Basal area of conspecific trees $\geq 10$ cm DBH ( $m^2$ ) |          |



### 3.4. Results

We first built the ZIP model without spatial effect: we selected variables  $\mathbf{B}$  (see eq. 6) using a logistic GLM of sapling presence/absence as proposed by Barry and Welsh (2002). Given  $\mathbf{B}$ , we then selected variables  $\mathbf{X}$  (see eq. 7) in a complete ZIP model. Covariables were selected among candidates variables (not shown) with a classical stepwise selection using Maximum Likelihood Estimation and Akaike Information Criterion (AIC, McCullagh and Nelder, 1989). The retained models were (see eq. 6 and 7)

$$\mathbf{B}\gamma_{\mathbf{s}} = Int_1 + \gamma_{\mathbf{s}}^1 \mathbf{Ele} + \gamma_{\mathbf{s}}^2 \mathbf{Slo} + \gamma_{\mathbf{s}}^3 \mathbf{Re} + \gamma_{\mathbf{s}}^4 \mathbf{dna}$$

and

$$\mathbf{X}\beta_{\mathbf{s}} = Int_2 + \beta_{\mathbf{s}}^1 \mathbf{Ele} + \beta_{\mathbf{s}}^2 \mathbf{Slo} + \beta_{\mathbf{s}}^3 \mathbf{tFL} + \beta_{\mathbf{s}}^4 \mathbf{sD} + \beta_{\mathbf{s}}^5 \mathbf{G}_{\text{cons}} + \beta_{\mathbf{s}}^6 \mathbf{Re}$$

where  $Int_1$  and  $Int_2$  were two unknown intercepts and  $\gamma_{\mathbf{s}}$  and  $\beta_{\mathbf{s}}$  the unknown regression parameters. The same covariables were used for the ZINB and SZINB models. For the Poisson and Spatial Poisson models, we used covariables associated to the Poisson intensity model only,  $\mathbf{X}$  calibrated on the ZIP model.

Table 2 shows the Deviance Information Criterion associated with the eight models. The results underlined that each spatial version was better than the non spatial one. Moreover, taking into account zero inflation allowed to obtain better fits. While the NB model was better than the classical Poisson model, DIC associated to the ZINB model was higher than the DIC of the ZIP model. These results tend to underline that after taking into account the zero over-representation, data was not overdispersed. DIC value of the spatial Poisson showed that the model clearly improved the non spatial version of Poisson and Negative Binomial. The spatial effect could account for zero inflation by taking into account the spatial structure of the data. Finally, the SZIP model was the best one according to DIC. In the following, we present results only for the SZIP model.

TABLE 2. — Deviance Information Criterion of the eight estimated models.

| P      | NB     | ZIP    | ZINB   | SP     | SNB    | SZIP          | SZINB  |
|--------|--------|--------|--------|--------|--------|---------------|--------|
| 3772.4 | 3113.6 | 2722.9 | 2958.7 | 2662.5 | 3105.7 | <b>2527.3</b> | 2906.8 |

Table 3 presents the Posterior means and standard deviations of regression parameters. It appeared that distance to the nearest adult ( $dna$ ) had a strong influence on the Bernoulli part of the model. This result was consistent with the dispersal mode of the species, *i.e.* autochory, which implies short dispersal distances (most seeds land under the tree crown), even if secondary dispersal by water (hydrochory) may occur and increase those distances. The species

is abundant in the bottomlands (permanently waterlogged areas) of Paracou which may favor such dispersal. This topographic position was detected in the model because the presence of juveniles was negatively correlated with elevation (*Ele*). Slope (*Slo*) and recruitment (*Re*) had non significant effect on the Bernoulli distribution: this is coherent with the mid-tolerant temperament reported for the species. Likewise, no specific variable had clear influence on the intensity of the Poisson distribution. However, the loss of basal area in standing deaths (*sD*), and recruitment (*Re*) to a lesser extent, had non-zero effects (Table 3). Hence, juveniles could take advantage of the standing deaths of large trees to durably settle in their neighborhood.

Weak effects of covariables on the intensity of the Poisson distribution were due to the strong influence of the spatial process. Indeed, in the non-spatial ZIP covariables were initially selected because of their explanatory power on juveniles intensity. In the spatial version, they finally showed non-significant or weak effects. The spatial correlation ( $\rho$ ) and the precision of the CAR model ( $1/\sigma$ ) were estimated to 0.984 (sd: 0.0112) and 2.899 (sd: 0.4522). These results underlined the strong autocorrelation in the observed density. The variogram of models residuals (Fig. 1) shows that the CAR distribution fairly took spatial dependency into account as a prior distribution for the spatial process.

TABLE 3. — Posterior mean and standard deviation (sd) of regression parameters in the Bernoulli ( $\omega$ ) and Poisson ( $\lambda$ ) intensities in the SZIP model.

|                         | $\omega$ |         | $\lambda$               |                  |
|-------------------------|----------|---------|-------------------------|------------------|
|                         | mean     | sd      | mean                    | sd               |
| <i>Int</i> <sub>1</sub> | 5.34900  | 0.80720 | <i>Int</i> <sub>2</sub> | -0.44150 0.10550 |
| <i>Ele</i>              | -1.84900 | 0.52400 | <i>Ele</i>              | -0.12440 0.12700 |
| <i>Slo</i>              | -0.03705 | 0.39700 | <i>Slo</i>              | 0.05044 0.06599  |
| <i>Re</i>               | -0.31240 | 0.33050 | <i>tFL</i>              | 0.05310 0.08128  |
| <i>dna</i>              | 13.52000 | 2.35300 | <i>sD</i>               | 0.12860 0.06292  |
|                         |          |         | <i>G<sub>cons</sub></i> | -0.07920 0.07533 |
|                         |          |         | <i>Re</i>               | 0.14710 0.12400  |

Finally, figure 2 presents the observed and predicted spatial juvenile patterns in the four permanent sample plots of the study. The two patterns showed good agreement which underlined the efficiency of the SZIP model to take into account the features of the observed pattern, that are zero-inflation and autocorrelation, through the HB approach.

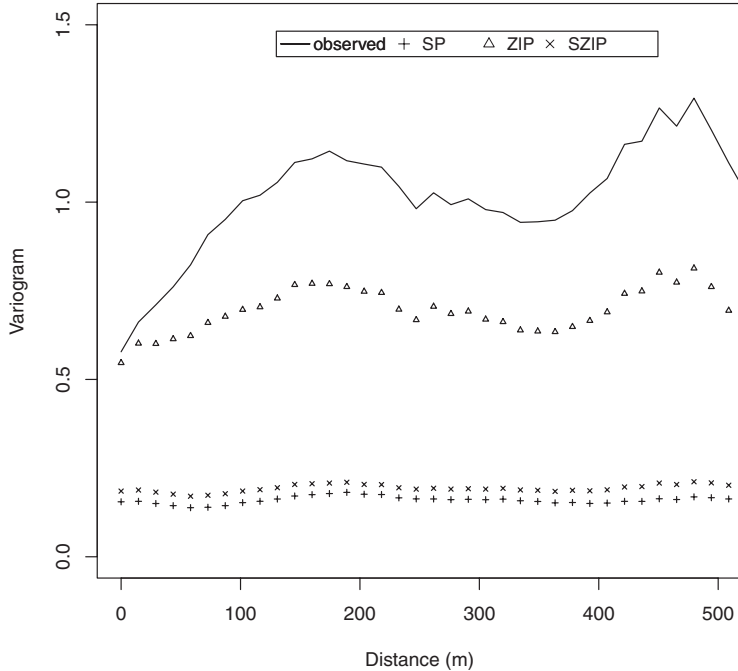
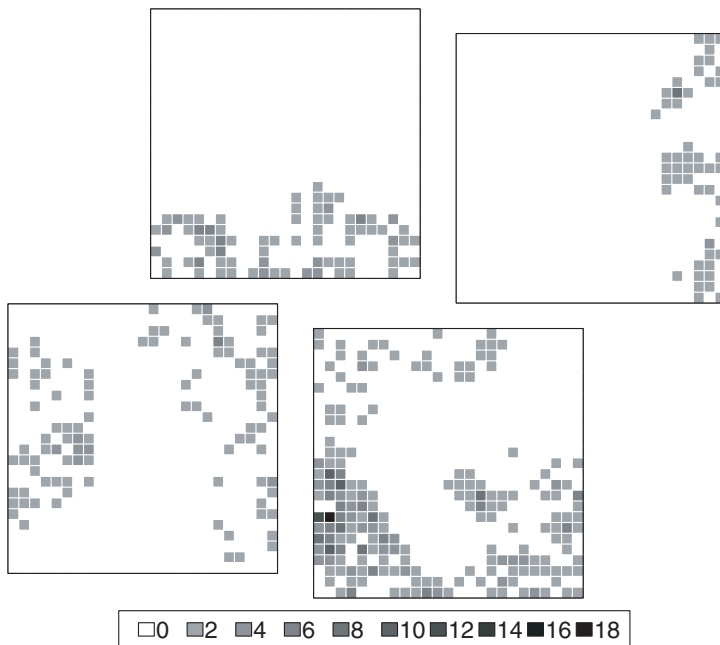


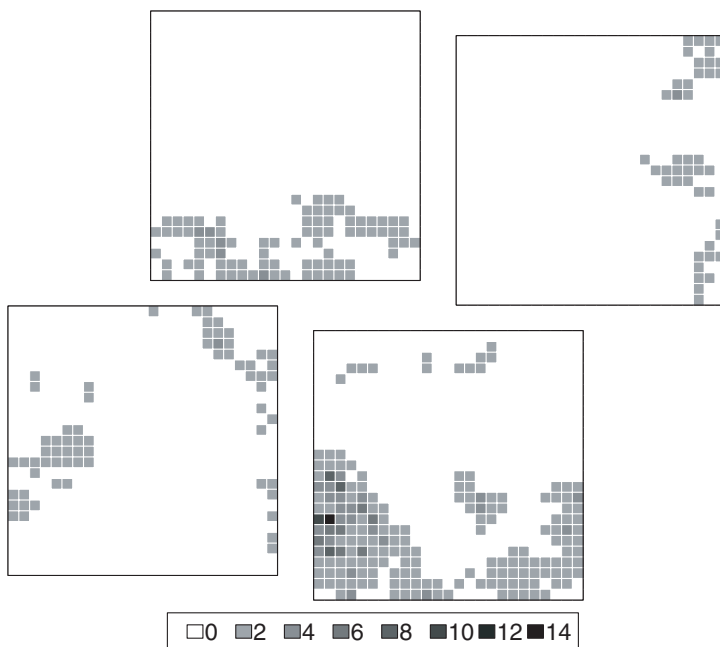
FIG 1. — Spatial structure at the site scale: the solid line shows the empirical variogram of juvenile density (observed), while symbols show variograms calculated on the residuals of the three models (Spatial Poisson, SP; Zero-Inflated Poisson, ZIP; Spatial ZIP model, SZIP).

#### 4. Conclusion

This study underlined the relevance of the Hierarchical Bayesian approach to model spatial patterns of early-life stages in plant populations. The HB framework allowed to take random spatial effects into account, which is critical in highly heterogeneous ecosystems such as tropical forests. The flexibility of HB models could allow to include more sources of uncertainty, or other effects regarding dispersal. But Markov *prior* specification for the spatial effect appeared to be very strong. When the random spatial effect was included, the effects of the main covariables strongly decreased as if they were hidden by the CAR structure. Consequently, a more realistic treatment of spatial patterns would certainly refine our understanding of disturbance and dispersal effects on tropical tree species populations and increase the predictive power of models.



(a)



(b)

FIG 2. — Observed (a) and estimated (b) spatial patterns of *E. falcata*: local density of juveniles between 1 and 4 cm DBH in 10 m×10 m cells.

## References

- BANERJEE S., CARLIN B. and GELFAND A. (2003). *Hierarchical modeling and analysis for spatial data*. Monographs on Statistics and Applied Probability 101. Chapman Hall/CRC.
- BARRY S. and WELSH A. (2002). Generalized additive modelling and zero inflated count data. *Ecological Modelling*, **157(2-3)**, 179-188.
- BESAG J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B*, **36**, 192-236.
- CELEUX G., FORBES F., ROBERT C. P. and TITTERINGTON D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis (in press)*.
- CLARK J. (2005). Why environmental scientists are becoming bayesians? *Ecology Letters*, **8**, 2-14.
- COLLINET F. (1997). *Essai de regroupement des principales espèces structurantes d'une forêt dense humide d'après l'analyse de leur répartition spatiale (Forêt de Paracou - Guyane)*. Ph.D. thesis, Université Claude Bernard - Lyon 1.
- CONDIT R., ASHTON P., BAKER P., BUNYAVEJCHEWIN S., GUNATILLEKE S., GUNATILLEKE N., HUBBELL S., FOSTER R., ITOH A., LAFRANKIE J., LEE H., LOSOS E., MANOKARAN N., SUKUMAR R. and YAMAKURA T. (2000). Spatial patterns in the distribution of tropical tree species. *Science*, **288**, 1414-1418.
- FAVRICHON V. (1994). Classification des espèces arborées en groupes fonctionnels en vue de la réalisation d'un modèle de dynamique de peuplement en forêt guyanaise. *Revue Ecologie-La Terre et La Vie*, **49**, 379-403.
- FLORES O., GOURLET-FLEURY S. and PICARD N. (2006). Local disturbance, forest structure and dispersal effects on sapling distribution of light-demanding and shade-tolerant species in a french guianan forest. *Acta Oecologica*, **29(2)**, 141-154.
- GOURLET-FLEURY S., GUEHL J. and LAROUSSINIE O. (2004) *Ecology and Management of a Neotropical Rainforest - Lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Elsevier, Paris.
- LEGENDRE P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74(6)**, 1659-1673.
- MCCULLAGH P. and NELDER J. (1989) *Generalized Linear Models*. Monographs on Statistics and Applied Probability 37. London, Chapman & Hall edition.
- OAB-OIBT (2003). *Principes, critères et indicateurs OAB-OIBT de la gestion durable des forêts tropicales naturelles d'Afrique*. Série Développement de politiques OIBT 14, Organisation africaine du bois (OAB), Libreville, Gabon et Organisation internationale des bois tropicaux (OIBT), Yokohama, Japon.
- PLACKETT R. (1981). *The analysis of categorical data*. Griffin.
- R DEVELOPMENT CORE TEAM (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.
- RIDOUT M., DEMETRIO C. and HINDE J. (1998). Models for count data with many zeros. In *International Biometric Conference*. Cape Town. 14 Bayesian models of tree density
- SPIEGELHALTER D., BEST N., CARLIN B. and VAN der LINDE A. (2002). Bayesian measures of model complexity and t (with discussion). *Journal of the Royal Statistical Society, Series B*, **6**, 583-639.
- WIKLE C. (2003). Hierarchical bayesian models for predicting the spread of ecological processes. *Ecology*, **84**, 1382-1394.