

Hierarchical Bayesian models accounting for spatial dependence and zero-inflation of sapling pattern in French Guianian forest

Flores O. ^{a,*}, Mortier F., ^b Dessard H., ^a Gourlet-Fleury S. ^a

^a*CIRAD - UPR Dynamique forestière, TA 10/D, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France*

^b*CIRAD - UPR Génétique forestière, TA 10/C, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France*

Abstract

1. Modelling the spatial patterns of trees in early life-stages provides insights into driving ecological processes and species realized niches. Still few studies address challenging features of spatial data which may mislead ecological inference. This paper presents Hierarchical Bayesian (HB) models of sapling abundance in a tropical forest handling spatial correlation and zero inflation in the studied patterns.

2. We questioned how distributions of saplings related with physical conditions, disturbance, stand structure and dispersal around adults for 6 tropical tree species with different biological attributes: *Oxandra asbeckii*, *Eperua falcata*, *Eperua grandiflora*, *Dicorynia guianensis*, *Qualea rosea*, *Tachigali melinonii*. The study was conducted in the Paracou experimental site, French Guiana, where stands experienced silvicultural treatments in permanent sample plots from 1986 to 1988. Qualitative and quantitative variables obtained from census data and GIS layers described local ecological conditions.

3. Three types of models were built through a Hierarchical Bayesian approach: spatial generalized linear mixed models (SGLM), zero inflated Poisson models (ZIP), spatial zero inflated mixed Poisson models (SZIMP). Spatial dependence of sapling pattern was modeled through a Conditional Auto Regressive process (CAR).

4. Comparison of the models showed that sapling abundance was better explained when autocorrelation was taken into account. In SGLM, the spatial process alone could correct zero-inflation influence on fits, while in SZIMP models, the links between ecological descriptors and the response could take various forms. This work emphasized the relevance of HB models handling autocorrelation to improve modelling of abundance-environment relationships.

5. Effects of ecological variables differed between the spatial and non-spatial cases, and therefore depended on statistical hypotheses. Species responses differed along

two gradients of disturbance and topography, as well as the dispersal patterns of saplings around adults. Findings were not systematically consistent with an *a priori* knowledge of species shade-tolerance and seed dispersal modes, and thus rose new hypotheses.

Key words: conditional autoregressive model, disturbance, French Guiana, hierarchical bayesian models, sapling pattern, spatial dependence, zero inflated Poisson, Paracou, tropical rainforest.

1 Introduction

At local scales, a large variety of processes, biotic and abiotic (e.g. dispersal, competition, disturbance (Molino and Sabatier, 2001)) and physical conditions (e.g. topography, He *et al.*, 1997; Plotkin *et al.*, 2002; Valencia *et al.*, 2004, and waterlogging Pélissier and Goreaud, 2001; Harms *et al.*, 2001) interact to control species distributions. The study of those patterns can theoretically provide insights into species realized niches and driving ecological processes (Austin, 2002), in particular into the respective part of dispersal (dispersal limitation, e.g. Dalling and Wirth (1998)) and environmental conditions (niche theory, e.g. Grubb (1977)). Today, a growing quantity of detailed ecological data are available on experimental sites. Methods, based on information-theoretic approaches such as model selection, allow to systematically explore and compare relations between ecological processes and the local abundances of species (Rushton *et al.*, 2004).

Statistical modelling studies rely on three components (Austin, 2002), at least implicitly: an *ecological model*, which addresses hypotheses or concepts in a given context, a *data model*, which assumes relations between sampled data and focused ecological processes at a given scale, and finally a *statistical model* linking responses and ecological variables of the data model. Assumptions made for one model may have implications on the other models, and therefore influence ecological conclusions (Austin, 2002; Keitt *et al.*, 2002). Following this three-components frame, we present here a comparative approach to model sapling abundance of six tropical tree species with different biological attributes at intermediate scale (*ca* 30 ha).

The individualistic perspective of communities defines a relevant *ecological model* for spatial modelling (Guisan and Zimmermann, 2000) in which

* Corresponding author.

Email address: flores@cirad.fr (Flores O.).

species interact with their environment through intrinsic rules. We considered here an individualistic ecological model in which trees occur at a given stage depending on topography and soil conditions (waterlogging), biotic interactions (through dispersal and competition) and disturbance. Among ecological processes affecting species patterns, disturbance is of outmost importance as it controls the local dynamics of trees populations (Shugart, 1984; Sheil, 1999) through the release of competition in treefall gaps. Early-life stages patterns may then reflect past patterns of disturbance within the stands (Nicotra *et al.*, 1999). Special attention was devoted here to include disturbance as a possible determinant of species patterns.

Tropical trees species exhibit large differences in growth potential and spend variable periods of time in early life-stages (Clark and Clark, 1999). Effects of past disturbance events are therefore likely to be perceptible in size-classes differing between focal species (Molino and Sabatier, 2001). Moreover, describing the environment of trees in tropical forests remains a difficult task, partly because of high spatial heterogeneity (Ricklefs, 1977). To allow comparison among the studied species, our *data model* defined specific DBH classes for the sapling stage. Abundances were evaluated into sampling cells on an exhaustive and regular basis in permanent sample plots (PSP) of French Guiana. Indirect ecological variables described the processes of the ecological model on the same locations. Such indirect variables constitute proxies of direct (physiological) or resource gradients (Guisan and Zimmermann, 2000).

Spatial patterns of tropical tree species are often clumped (Condit *et al.*, 2000), so that abundance data are likely to be spatially correlated, or autocorrelated (Legendre, 1993). Autocorrelation challenges the common statistical hypothesis of observations being independent. Quadrat-sampling may also induce over-dispersed data due to many zero counts, or *zero inflation* (McCullagh and Nelder, 1989; Ridout *et al.*, 1998). In this paper, we specify *statistical models* to handle autocorrelation and zero-inflation through a Hierarchical Bayesian approach (HB, Clark (2005)). This approach allows to model complex biological data into a series of simpler conditional models (Wikle, 2003; Clark, 2005).

Within this framework, we address a double question: what are the main determinants of sapling pattern of the six species among physical conditions, disturbance, stand structure and intraspecific relations ? Do models taking autocorrelation into account better explain sapling pattern than simpler models and in which way ?

2 Material and Methods

2.1 Study site

The study was conducted at the Paracou experimental site (5°18' N, 52°23'W) in a *terra firme* rain forest. The site lies in the coastal part of French Guiana in an equatorial climate with two main seasons. A dry season occurs from August to Mid-November. From March to April, a short drier period interrupts the rainy season. Annual rainfall in the vicinity of the site is 3041 mm (Gourlet-Fleury *et al.*, 2004). At Paracou, streams incise a smooth geomorphological system mainly on shallow ferralitic soils. Part of the site is covered by permanently waterlogged areas with particular floristic composition.

The design of the site consists in twelve 300 × 300 m permanent sample plots with a 25 m inner buffer zone. In each central 250 × 250 m square, all trees ≥ 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. Three treatments were applied during the 1986-1988 period combining selective logging of increasing intensity and additional poison-girdling. Major interests of the Paracou experimental site are the wide range of disturbance experienced by the stands, and the 20-years long monitoring of trees ≥ 10 cm DBH.

The present work concerns four adjacent permanent sample plots gathering an undisturbed control plot and three treated plots, between 10 and 42 m above sea-level. Two periods were distinguished in the recent plots history: the *logging* period from 1986 to 1988 and the *recovery* period from 1989 to 2002 (Gourlet-Fleury *et al.*, 2004).

2.2 Focal species, life-stages and response variables

Commonness, light requirement and dispersal modes were criterion of species choice (Sabatier, 1983; Favrichon, 1995; Gourlet-Fleury and Houllier, 2000). The six focal species were, by growing order of light-requirement: one sciaphilous species *Oxandra asbeckii* Pulle, R.E.Fr. (*Annonaceae*), three tolerant to mid-tolerant species (*Eperua falcata* Aublet, *Caesalpinaceae*, *Eperua grandiflora* Aublet, Benth., *Caesalpinaceae*, *Dicorynia guianensis* Amshoff, *Caesalpinaceae*), two light-demanding species (*Qualea rosea* Aublet, *Vochysiaceae*, *Tachigali melinonii* Harms, *Caesalpinaceae*). *O.asbeckii* is an endozoochorous species of the understory, the highest trees staying below 15 m high. *E.falcata* is an autochorous species and *E.grandiflora* a barochorous species

that both occur in the canopy at maximal heights of 30-35 m (Sabatier, 1983). *D.guianensis*, *Q.rosea* and *T.melinonii* are anemochorous species of the canopy with emergent trees reaching 40 meters (Favrichon, 1995). *T.melinonii* is the fastest-growing and *a priori* most light-demanding species of the study.

In 2002-2003, all plants with $1 \text{ cm} \leq dbh \leq 10 \text{ cm}$ were sampled and georeferenced. DBH were recorded in 1 cm classes. For the present study, we restricted the *sapling stage* to plants whose installation occurred most likely during the post-logging period. For each species, the sapling stage was thus limited by a specific upper DBH limit (D_{sap}) accounting for differences in average growth among species (Gourlet-Fleury, unpublished data). Thus, DBH classes for saplings were [1 – 2] cm for *O.asbeckii*, [1 – 3] cm for *E.grandiflora*, [1 – 4] cm for *E.falcata*, [1 – 5] cm for *D.guianensis*, [1 – 6] cm for *Q.rosea* and [1 – 9] cm for *T.melinonii*. Sapling abundance which served as response variables in the statistical models was then calculated in $10 \times 10 \text{ m}$ cells (625 within each PSP).

The adult stage gathered potential mother-trees over a given DBH at maturity derived from literature (Doligez, 1996) or defined regarding the status of the species (Collinet, 1997). DBH at maturity was 10 cm for *O.asbeckii*, 25 cm for *D.guianensis*, 35 cm for *E.falcata*, *E.grandiflora*, *Q.rosea* and *T.melinonii*. Potential mother-trees included naturally dead and logged trees for each species.

2.3 Ecological descriptors and gradients

Ecological variables derived either from available GIS maps of the study site or from census data of trees $\geq 10 \text{ cm}$ DBH. Topography (elevation and slope) derived from a Digital Elevation Model (DEM) of the site. Three qualitative variables characterized the position of sampling cells regarding seasonal streams and bottomlands, skid trails and logging damage (Table 1). Static and differential quantitative stand variables were calculated using basal areas on 20 m-radius circular subplots centered on the sampling cells. Differential variables concerned either the *logging* period, or the *recovery* period.

Static variables described the local forest structure in 2002 and concerned the total basal area, the basal area of pioneer taxa and the first two axes of a CA (Correspondence Analysis) on diameter distributions (see Table 1). Among differential variables, disturbance variables separately quantified trees deaths in treefalls or standing deaths (Table 1). Two variables quantified trees recruitment over 10 cm DBH and the gross change in basal area during the recovery period. Mean and standard deviation of treefalls ages characterized the temporal pattern of local disturbance during the recovery period.

Finally, three *population* variables characterized *intraspecific* interactions between saplings and surrounding conspecific trees (Table 1): the distance from cells centers to the nearest adult accounted for dispersal, the basal area of conspecific trees (≥ 10 cm DBH) in 2002 accounted for intraspecific competition, the loss of basal area from conspecific trees (≥ 10 cm DBH) during the recovery period accounted for a possible release of competition through conspecific deaths. For the last two variables, a 20 m-radius was used as for other stand variables.

For the sake of clarity in model interpretation :

- (1) we considered population variables separately from other variables,
- (2) we summarized environmental heterogeneity along two gradients formed by combinations of the environmental variables.

These gradients served to predict species response according to best models predictions. They were build from the first two axes of a PCA on environmental variables. The first axis was positively supported by variables LD , M_{tfL} , $Recru$, $diam1$ and G_{pio} (24% of inertia explained, Table 1). Hence, it indicated a gradient of *logging disturbance* during the *logging* period. The second axis was positively supported by Ele , M_{tfR} and A_{tfR} , and negatively by WL and dG_R (inertia explained: 11%). The second gradient reflected *topographic position*, and possible disturbance during the *recovery* period. Local conditions along the second axis varied from undisturbed subplots near bottomlands ($WL = 2$) to subplots on plateaux possibly disturbed by treefalls during the recovery period. We characterized one gradient along each of the two axes by sampling cells according to their scores and their low inertia on other axes of the analysis ($n_{logging} = 37$, $n_{topo} = 42$).

2.4 Statistical models

A Hierarchical Bayesian approach. Hierarchical Bayesian modelling aims at decomposing a complex problem into a series of simpler conditional levels (Banerjee *et al.*, 2003; Wikle, 2003): at a given hypothesis level, inference conditionally rely on hypotheses made at higher levels. In the following section, we present spatial Poisson models in the HB context. We then present Zero Inflated Poisson models (ZIP) and extend the ZIP formulation to include autocorrelation. Finally, we focus on model calibration and evaluation to address the quality of fits and variables effects.

Spatial Poisson models. Abundances are classically modelled using Poisson distributions and the *log* function to link the Poisson intensity λ to ecologi-

cal variables (McCullagh and Nelder, 1989). A major issue of spatial modelling is to correctly describe the covariance structure of the data, i.e. possible autocorrelation in observations. The idea here is to include a spatial effect $\alpha(s)$ that accounts for spatial dependence in local Poisson intensities. $\alpha(s)$ is modelled as a random field over discretized space, and $Z(s)$ is the count of saplings in the cell at location s .

In the HB context, three basic levels constitute a model: a *data* level, specifying the conditional distribution of the data Z given parameters and underlying processes, a *process* level specifying the conditional distributions of the processes given their own parameters, and a *parameter* level specifying prior distributions for all parameters (Wikle, 2003):

$$\begin{aligned} \text{data level: } Z(s)|\lambda(s) &\sim \text{Poisson}(\lambda(s)) \\ \text{process level: } \log(\lambda(s))|\mu, \delta, \alpha(s) &= \mu + \mathbf{P}\delta + \alpha(s) \\ \text{parameter level: } &\text{prior distributions of } \mu, \delta \text{ and parameters for } \alpha, \end{aligned} \tag{1}$$

where μ is an intercept, \mathbf{P} a matrix of ecological variables, δ a vector of regression parameters, s a vector of spatial locations $\alpha(s)$ a spatial random effect. Given parameter $\lambda(s)$, *conditional* independence between observations replaces the usual hypothesis of complete independence at the data level. The purpose of the Bayesian analysis is then to estimate the conditional posterior distribution of the parameters given data and processes.

Conditional Autoregressive model (CAR). We retained a Conditional Auto-Regressive model for $\alpha(s)$ to take correlation between neighboring observations into account (CAR, Besag (1974)). For each cell, we used a Moore neighborhood (the chess king's move). The spatial process intensity, $\alpha(s)$, followed a *conditional* gaussian distribution given intensities in the neighborhood:

$$\alpha(s_i)|\alpha(s_j), j \in v_i \sim \mathcal{N}\left(\rho \sum_{j \in v_i} w_{ij} \alpha(s_j), 1/\tau\right), \tag{2}$$

where ρ and τ are two unknown parameters, (w_{ij}) a set of known spatial weights and v_i the neighborhood of s_i (Banerjee *et al.*, 2003). ρ measures the strength of the relation between $\alpha(s_i)$ and v_i . τ is the conditional precision of the process ($1/\tau$ is the conditional variance).

Zero Inflated models. Because count data are often over-dispersed due to over-represented zero counts (Ridout *et al.*, 1998), fits with Poisson distributions may be poor. We modelled zero-inflation of sapling abundance with a

special case of finite mixture models, i.e. Zero Inflated Poisson (ZIP) models (Lambert, 1992). In the ZIP scheme, data proceed from a two-stages regime (Zorn, 1996). In a first – *transition* – stage, the outcome of a Bernoulli process determines the intensity of the second stage, which is then either strictly nul or not. In the second – *events* – stage, a Poisson process of the given intensity determines the final observation. Hence, the response variable Z can be modelled as: $Z = B(\omega)P(\lambda)$, where B is a Bernoulli random variable indicating saplings *absence* with probability ω ($B = 1$ implies $Z = 0$), and P a Poisson random variable describing sapling abundance with intensity λ . The data distribution \mathcal{Z} is then a mixture of two Poisson distributions: $\mathcal{Z} = \omega\mathcal{P}(0) + (1 - \omega)\mathcal{P}(\lambda)$, where $\mathcal{P}(0)$ is the zero-point probability mass function and ω the unknown proportion of mixture between the two distributions. Thus, observations in Z proceed either from a null distribution or from a classical Poisson distribution. A major interest of ZIP models is that parameters ω and λ can rely on different set of variables. Truncated ZIP models, or Hurdle models, separately model presence-absence with a Bernoulli distribution and non zero counts with a truncated Poisson distribution (Welsh *et al.*, 1996; Bar-Hen, 2002). We preferred the mixture specification because zero counts could arise either from the Bernoulli or from the Poisson distribution.

Spatial Zero Inflated Mixed models. Following the HB approach, we extended the ZIP formulation to include autocorrelation. At the *data level*, we supposed that the response variable Z was spatialized, $Z = Z(s)$ and ZIP distributed: $Z(s)|\omega(s), \lambda(s) \sim \mathcal{Z}(\omega(s), \lambda(s))$. Given the two main parameters $\lambda(s)$ and $\omega(s)$, observations in Z were assumed to be *conditionally* independent.

At the *process level*, $\omega(s)$ and $\lambda(s)$ were linked with ecological variables through canonical link functions and we defined: $u(s) = \text{logit}(\omega(s))$ and $v(s) = \log(\lambda(s))$ (McCullagh and Nelder, 1989). For simplicity, we included explicit spatial dependence among observations only in the Poisson part of the ZIP model (see also Wikle and Anderson (2003)). Thus, the Poisson process intensity $\lambda(s)$ depended on both a set of specific variables and an underlying spatial process $\alpha(s)$. As in the Spatial Poisson case, $\alpha(s)$ followed a CAR model:

$$u(s)|\mu_1, \gamma = \mu_1 + \mathbf{B}\gamma \tag{3}$$

$$v(s)|\mu_2, \beta, \alpha(s) = \mu_2 + \mathbf{M}\beta + \alpha(s) \tag{4}$$

with (μ_1, μ_2) two intercepts and (\mathbf{B}, \mathbf{M}) two sets of selected variables.

Finally, at the *parameter level*, we used weak or non informative prior

distributions of parameters. For regression parameters γ and β , gaussian distributions were used. Prior for ρ was uniform on a constrained interval (see Banerjee *et al.* (2003) for details). Prior for τ was an inverse gamma distribution.

Model calibration and evaluation. For each species, we first built two models without spatial effect, a simple Poisson GLM and a ZIP model. In the ZIP case, we first selected variables \mathbf{B} using a logistic GLM of sapling presence/absence, as usually proposed for Hurdle models (Bar-Hen, 2002). Given \mathbf{B} , we then selected variables \mathbf{M} in a complete ZIP model. Variables were selected among candidates variables (Table 1) with a classical stepwise selection using Maximum Likelihood Estimation and Akaike Information Criterion (AIC) (McCullagh and Nelder, 1989).

Models were then recalibrated through the HB approach. Three models were retained per species: a Spatial Generalized Linear Mixed model (SGLM), a non-spatial Zero Inflated Poisson model (ZIP), and a Spatial Zero Inflated Mixed model (SZIMP). We discarded the non spatial Poisson GLM because of poor results. Model calibration was performed using WinBUGS software (Spiegelhalter, 2004) with 100000 iterations on one Monte Carlo Markov chain (MCMC) including an initial burning step of 20000 iterations. Other analyses were performed with R (R Development Core Team, 2004).

Model comparison in HB context is not a simple task. The *effective* number of parameters or degrees of freedom is not always clearly defined for hierarchical models (Spiegelhalter *et al.*, 2002) and can be very different from the actual number of parameters. 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 deviance moments to compare complex hierarchical models. DIC is defined as: $DIC = \overline{D(\theta)} + p_D$, where θ is the parameter set of the model, $\overline{D(\theta)}$ the mean of the Bayesian deviance $D(\theta)$ for all MCMC samples. p_D is the *effective* number of parameters and is defined as $p_D = \overline{D(\theta)} - D(\bar{\theta})$, where $\bar{\theta}$ is the mean of all MCMC samples of θ . p_D is proportional to the deviance variance and is regarded as a measure of the model complexity (Spiegelhalter *et al.*, 2002).

For a given variable, the mean of the posterior distribution defined the *variable effect*. In ZIP and SZIMP models, a variable was said *dual*, following Zorn (1996) terminology, if it occurred in both \mathbf{B} and \mathbf{M} matrices (Eq. 3 & 4). The response curve – the relation between the response and variable – to a dual variable, say x , depends on the ratio $\sigma_x = \frac{\beta_x}{\beta_x - \gamma_x}$ where β_x and γ_x are the

regression coefficients associated with x in \mathbf{B} and \mathbf{M} respectively. Depending on the sign and value of σ , the response curve can be of various shapes, either strictly monotonous or with a local maximum symmetric or not about this maximum (Fig. 5.a).

3 Results

3.1 Local abundance data and actual spatial patterns

Zero-inflation of sapling abundance varied between species with zero-frequencies between 58% for *O.asbeckii* and 87% for *T.melinonii*. *O.asbeckii* was the most common species regarding cells occupation (42%) and total saplings numbers (2271). On the opposite, *D.guianensis* and *T.melinonii* had the lowest total saplings numbers (615 and 616) and the lowest maximal abundances (8 and 11). *Q.rosea* was the most locally abundant species (34, total: 1197) and also the most variable in abundance. *E.falcata* and *E.grandiflora* occurred in 17% and 20% of the cells respectively with 17 and 11 saplings at maximal abundances (total: 807 and 861).

Moran's I (I_M) were calculated to estimate the sign and strength of local dependence between observations relatively to the total variance. We used here the neighborhood definition of the CAR model. All actual sapling pattern had positive I_M values (Fig. 3.a) with low variance ($<10^{-3}$, not shown) and thus were *positively* autocorrelated.

Empirical variograms characterized spatial patterns on the complete study area. Major changes in variograms slopes indicated spatial features (clumps) at various scales (Fig. 2, solid lines). A steep increase was observed up to about 50 m for *D.guianensis* and *E.grandiflora*, and up to about 100 m for *O.asbeckii*. For *E.grandiflora*, the variogram slowly decreased up to 500 m, which indicated that paired observations were mostly zeros at that distance. Variograms for *Q.rosea* and *E.falcata* showed a slow increase up to 200 m, with a larger range of variation for *Q.rosea*. For both species, variograms increased after 400 m, which indicated a separation of large clumps at this distance. For *T.melinonii*, the variogram showed a step increased in the first 30 m, but the species was the less structured in space.

3.2 Model evaluation

According to DIC values, SZIMP models were retained as best models for *D.guianensis*, *E.falcata*, *Q.rosea* *T.melinonii*, while SGLM were retained

for *O.asbeckii* and *E.grandiflora* (Table 2). For *O.asbeckii*, despite a weak difference in DIC between SGLM and SZIMP models, the SGLM was the best model because both the number of fixed regression parameters (33 *vs* 20) and the *effective* number of parameters (p_D) were lower for this model. Correlation coefficients showed good agreement between observations and fitted values. SGLM and SZIMP models showed similar Pearson and Spearman correlation coefficients (Table 2). Pearson coefficients were between 0.88 and 0.98 and Spearman coefficients were between 0.59 and 0.76.

For ZIP models, Pearson correlations coefficients were between 0.22 and 0.57 and Spearman correlation coefficients between 0.23 and 0.53. Large values of p_D indicated that model deviances varied widely along the Monte Carlo Markov Chains run for calibration. For *Q.rosea*, the poorer fit was essentially due to outliers, i.e. cells for which the ZIP model predicted high abundances whereas saplings were absent. Such behavior of the model will be discussed further.

Comparison of ZIP and SZIMP models showed that posterior means, thus variables effects, generally decreased when autocorrelation was taken into account, while standard deviations increased (Fig. 1).

3.3 Comparison of actual vs modelled patterns

Variograms of patterns modelled with mixed models closely paralleled actual sapling pattern, whereas ZIP models produced smooth or flat variograms (e.g. *D.guianensis*, Fig. 2). All modelled patterns showed a lower overall variance compared to observed distributions. This finding was confirmed with I_M since actual values were lower than values for calibrated patterns. I_M of mixed models (SGLM, SZIMP) were closer to actual values than ZIP values (Fig. 3.a). This ranking did not stand for *Q.rosea* which also showed the highest observed I_M .

I_M and the CAR precision parameter of mixed models (τ in Eq. 2) were positively related (Fig. 3.b). τ is inversely proportional to the variance of the spatial process α . Hence, the precision parameter τ strongly influenced the overall variability of modelled patterns (Fig. 3). Meanwhile, the ρ parameter showed little variability among models and had posterior means between 0.974 for *E.falcata* (SZIMP model, posterior standard deviation: $\sigma_\rho = 0.018$) and 0.999 for *O.asbeckii* (SGLM and SZIMP models, $\sigma_\rho = 0.001$). Such high values of ρ reveal strong positive autocorrelation and are consistent with I_M values about 0.5 (Banerjee *et al.*, 2003).

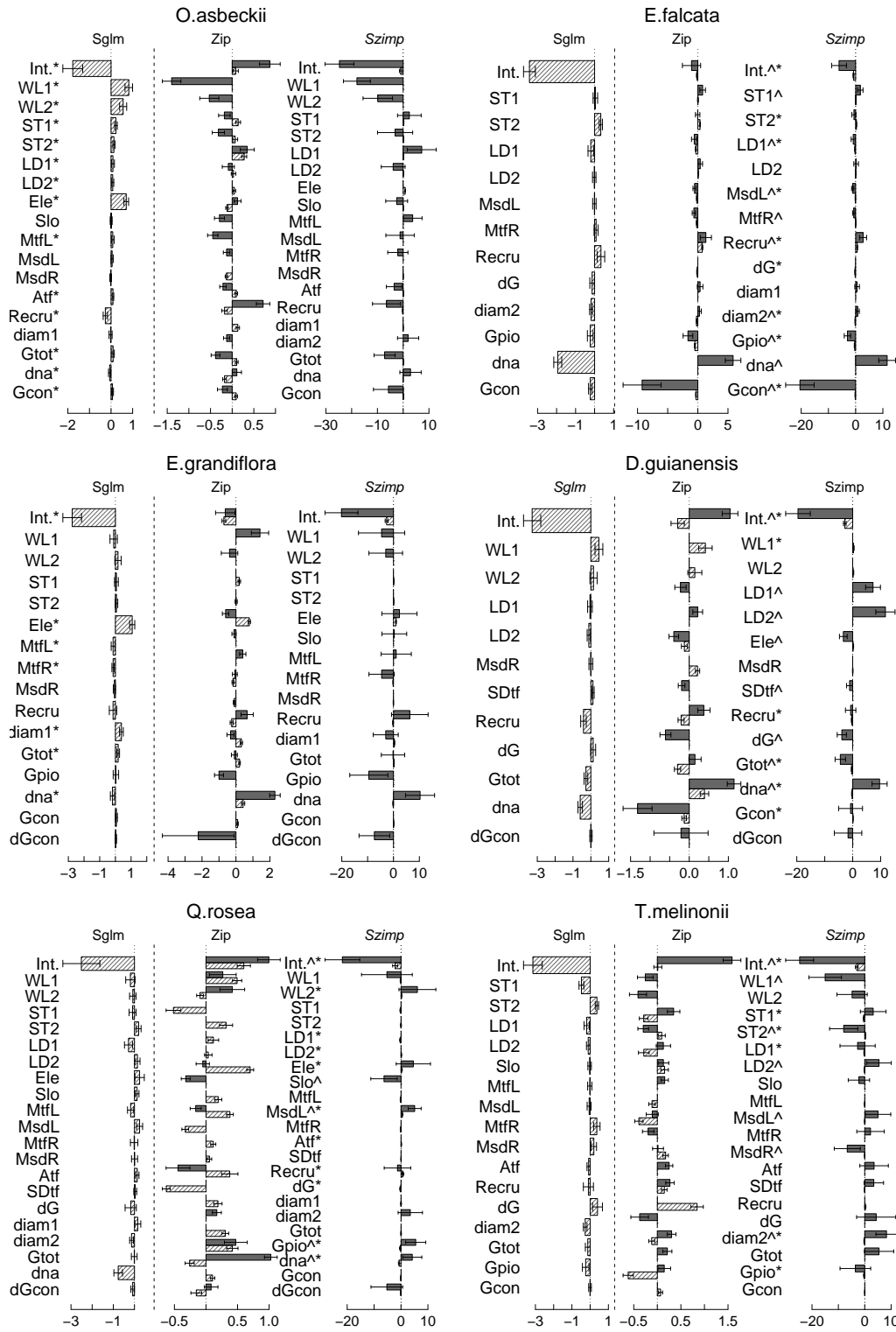


Fig. 1. Variables effects: posterior means and standard deviations intervals for fixed regression parameters, per species and per model type (see table 1 for labels, Int.: intercepts). For qualitative variables (LD, ST, WL), the effect in level 3 is: $\text{eff}_3 = -(\text{eff}_1 + \text{eff}_2)$. For best models (*italics*), symbols indicates intervals without 0 in. Shaded bars and * relate to coefficients of variables in matrix P or M, filled bars and ^ relate to coefficients of variables in matrix B.

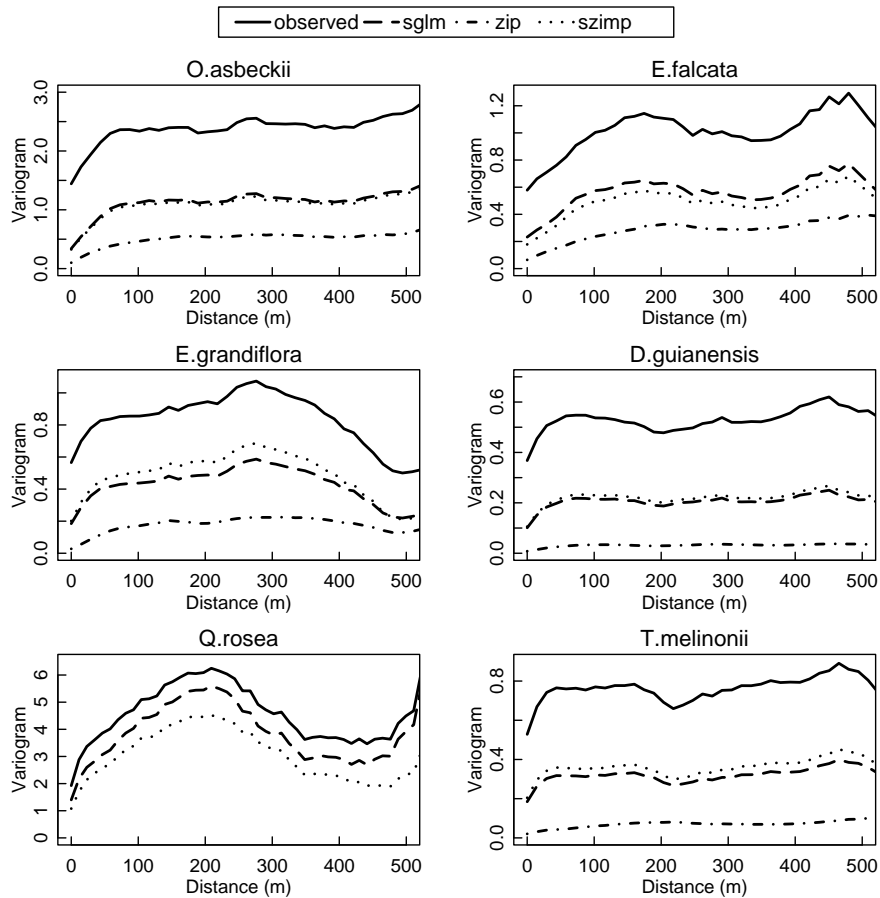


Fig. 2. Variograms of observed sapling abundance and modelled patterns calculated on the complete study area (the ZIP case is not shown for *Q.rosea* for graph clarity).

3.4 Adjusted sapling abundance along ecological gradients

Responses are presented here for the best selected models (Table 2). Along the gradient of *logging disturbance*, adjusted abundances for *O.asbeckii* decreased with increasing disturbance levels with constant variability. The response of *D.guianensis* was humped and maximal near the gradient middle (Fig. 4). Adjusted abundances of *E.falcata* and *Q.rosea* showed low mean and variance at low levels of *logging disturbance* (left end) and both increased with *logging disturbance*. *E.grandiflora* showed no particular trend in mean, but the variance increased at high levels of *logging disturbance*. *T.melinonii* showed the most marked positive trend along the disturbance gradient.

Along the gradient of *topography* (Fig. 4), sapling abundance increased with increasing scores for *D.guianensis*, *O.asbeckii*, *E.grandiflora*. Among variables most correlated with the second gradient (see section 2.3), changes in sapling abundance mostly related to elevation effects for *O.asbeckii* and

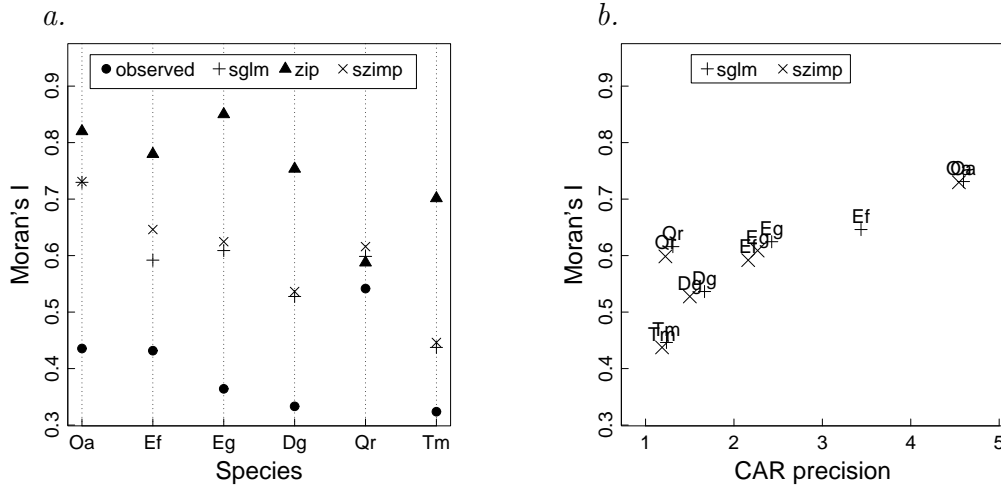


Fig. 3. *a.*: Comparison of Moran's I between actual and modelled sapling pattern for the six focal species. *b.*: Relations between Moran's I and the posterior CAR precision parameter (τ in Eq. 2) in the Spatial Mixed models (SGLM, SZIMP).

E.grandiflora (Ele on Fig. 1). Both Ele and dG had negative effects in the SZIMP model for *D.guianensis*. However, interpretation was not clear in this case because of opposed correlations signs of Ele and dG with the second gradient (Topo). For *Q.rosea*, sapling abundance weakly increased along the second gradient, while no trend was observed for *E.falcata* and *T.melinonii*. The variability increased along the gradient for *O.asbeckii* and *D.guianensis*, while no trend was visible for the other species.

3.5 Saplings dispersal patterns and relations with conspecific trees

Regarding intraspecific effects, the distance to nearest adult (**dna**) was the most informative population variables in 5 cases out of 6, while effects of those variables were low for *T.melinonii* (Fig. 1). For *E.falcata*, **dna** had a strong positive effect in **B** (Fig. 1), so that the response curve approached a logistic form with threshold at *ca* 50 m (Fig. 5.b). For that species, **G_{con}** was a dual variable (Fig. 1): sapling abundance increased with conspecific basal area up to and decreased over this threshold (Fig. 5.c). Sapling abundance of *O.asbeckii* decreased with increasing **dna** and increased with **G_{con}** (Figs. 1 and 5.b), but effects were relatively low. *O.asbeckii* had the lowest sampled distance range between cells and an adult (*ca* 0-50 m). On this range, the SGLM and SZIMP models produced similar response curves to **dna** (Fig. 5.b). For *D.guianensis*, **dna** was a dual variable: adjusted sapling abundance was nul above *ca* 65 m away from adult trees (Fig. 5.b). For *Q.rosea*, **dna** was dual but the adjusted response curve was strictly decreasing on the sampled range, *ca* 0-100 m (Fig.

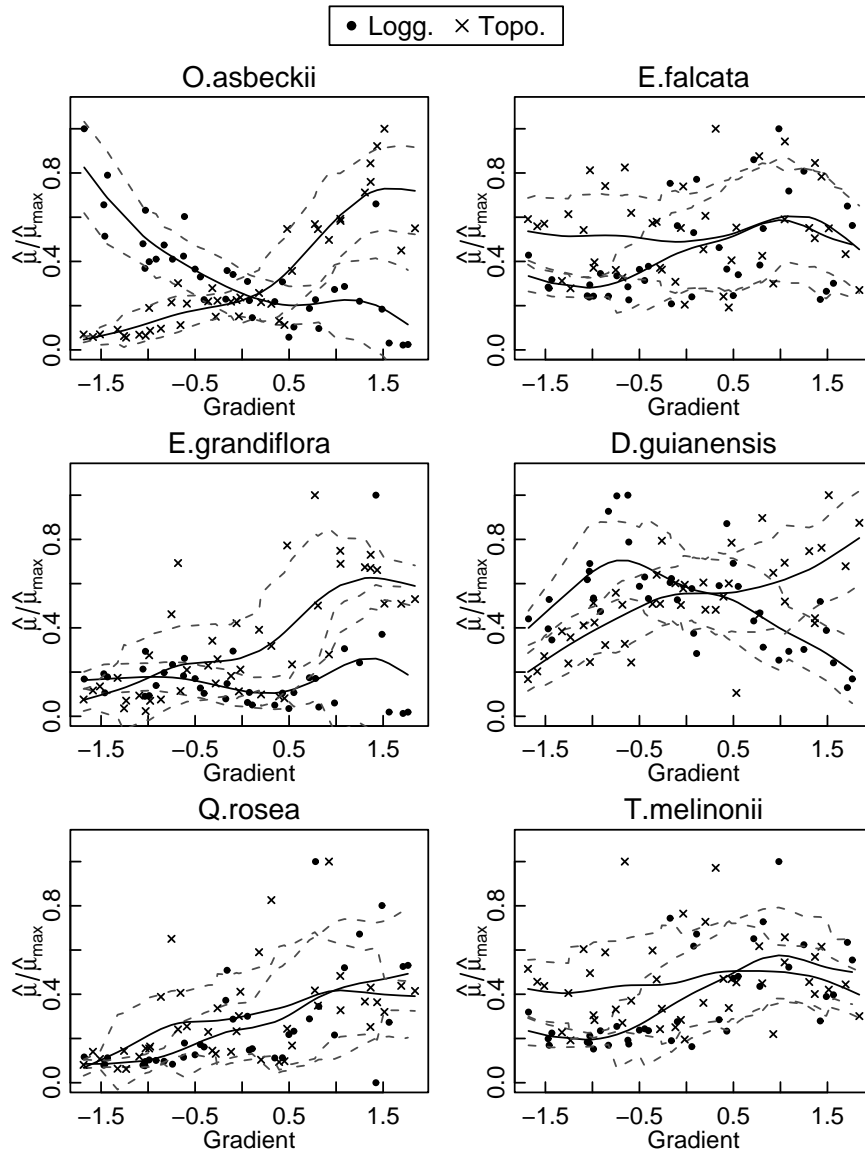


Fig. 4. Species responses along two ecological gradients. Dotted lines: mean of abundances, solid lines: intervals based on standard deviations. Logg.: gradient of logging disturbance (●), Topo.: gradient of topography from buffers around bottomlands to plateaux (×). *Y-axis*: sapling abundance predicted from best models normalized by maximum. See text for gradients construction.

5.b). The response curve of the best model for *E.grandiflora* (SGLM) decreased regularly up to 90 m, whereas in the SZIMP case, it showed a threshold at ca 60 m above which adjusted abundances were zero.

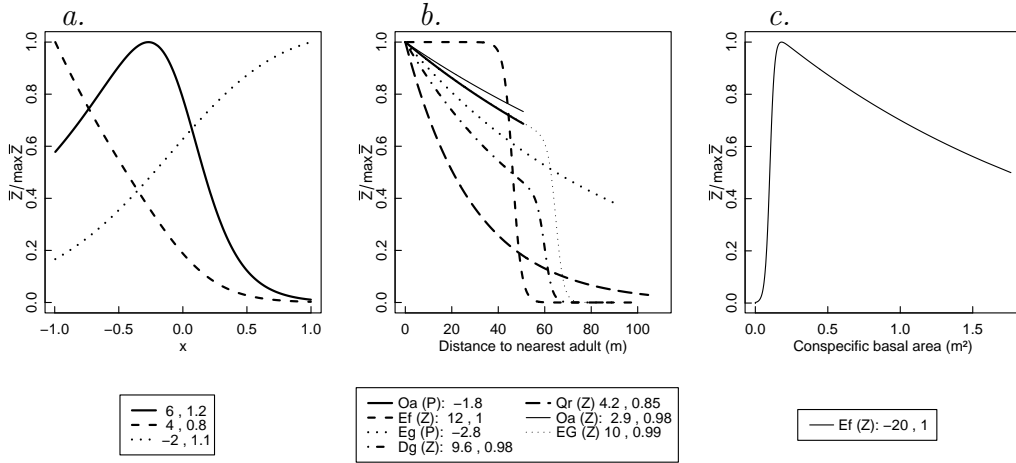


Fig. 5. Theoretical and adjusted responses curves. *X-axes*: explicative variables, *Y-axes*: normalized mean of abundances $Z: \frac{\bar{Z}}{\max(Z)}$.

Numbers are regression coefficients for variables on X-axis: for $Z \sim \text{ZIP}$: β, σ , with $\sigma = \frac{\beta}{\beta-\gamma}$ β, γ respectively Bernoulli and Poisson distributions coefficients for $Z \sim \text{Poisson}$: γ , Poisson distribution coefficient.

a.: theoretical responses curves to an uniform random variable on $[-1, 1]$ (x) for ZIP-distributed Z .

b.: adjusted response curves to distance to nearest adult for five species (observed ranges on X-axis). *Dg, Ef, Qr*: $Z \sim \text{ZIP}$, *Oa, Eg*, $Z \sim \text{Poisson}$. Dotted lines show the response curves for the SZIMP models for *Oa* and *Eg*.

c.: adjusted response curves to conspecific basal area for *Ef* with $Z \sim \text{ZIP}$.

Oa: *O.asbeckii*, *Ef*: *E.falcata*, *Eg*: *E.grandiflora*, *Dg*: *D.guianensis*, *Qr*: *Q.rosea*.

4 Discussion

4.1 Ecological aspects

Species susceptibility to disturbance and topography. We expected disturbance to induce a negative response of the sciaphilous species *O.asbeckii*, a positive response of the light-demanding species *Q.rosea* and *T.melinonii* and a humped response for the tolerant to mid-tolerant species *E.falcata*, *E.grandiflora* and *D.guianensis*. Regarding topography, *D.guianensis* and *E.grandiflora* are known to mainly settle on the upper part of slopes (Barthes, 1991; Bonjour, 1996) and *E.falcata* on bottomlands. Barthes (1991) showed that the two *Eperua* species are mutually exclusive in the stands. No a priori information was available for the three other species. A confusing effect between topography and small natural disturbances (treefalls) could occur along the topographical gradient, as both variables positively correlated to this gradient.

Species responses along the two gradients were coherent with our hypotheses for *D.guianensis* (disturbance + topography), *O.asbeckii*, *E.falcata*,

Q.rosea and *T.melinonii* (disturbance), *E.grandiflora* (topography). A positive effect of the topographical gradient was evidenced for *Q.rosea*, however difficult to interpret as the species is heliophilous and could react to small openings due to treefalls. On the contrary, the positive effect highlighted for *O.asbeckii* probably reflects a pure substrate effect as the species is sciaphilous. Surprises came from two particular features: (i) *E.grandiflora* appeared neutral on the disturbance gradient. This species finally appears less sensitive to disturbance and light than *E.falcata* and *D.guianensis*, as confirmed by a study of change in saplings stocks between 1992 and 1995 in Paracou plots (Rankin-de Mérona and Montpied, 2004), (ii) *E.falcata* appeared neutral on the topographical gradient. For this species, it seems that population variables were informative enough to mask effects of physical conditions because of strong clumping of saplings around potential mother-trees.

Saplings dispersal patterns and relations with conspecifics. Clumping among studied saplings populations could arise mainly because of limited dispersal around adults (Svenning, 2001), clumped seed dispersal (Howe, 1989; Russo and Augspurger, 2004), or survival response to patchy resource (Dalling and Wirth, 1998). Our results evidenced (i) a dispersal limitation of saplings for five species out of six, (ii) the persistence of seed dispersal patterns to some extent at the sapling stage despite mortality filters on earlier stages (Clark *et al.*, 1999; Wang and Smith, 2002) (iii) a lack of connection between saplings and adults patterns for the anemochorous and most light-demanding studied species *T.melinonii*.

The results obtained were coherent for the autochorous *E.falcata*, for which saplings were limited to low distances around adults. A similar pattern raised for the anemochorous *D.guianensis*. Despite wind dispersal, seeds and fruits reach maximal distances of 25 to 60 m around mother-trees (Sabatier, 1983; Loubry, 1993). Saplings appeared limited at the same scale. Saplings of *Q.rosea* were also mainly dispersed around mother-trees. But, despite similar dispersal mode and seed size compared to *D.guianensis* (Sabatier, 1983), *Q.rosea* was less limited. Differences in species fecundity could explain different dispersal abilities among these species (Clark and Ji, 1995). For the endozoochorous *O.asbeckii*, the abundance curve decreases slowly with distances between saplings and adults. Despite seeds dispersal by birds, median seed dispersal distances were estimated below 10 m in disturbed and undisturbed plots of Guyana (Ulft, 2004). Saplings-adults distances did not exceed 50 m here for this common species, which thus does not appear dispersal limited at the study scale.

The case of *E.grandiflora* can be questioned. The flexible SZIMP model

reveals a pattern similar to that of *E.falcata*, coherent with heavy seeds dispersed close to the mother-trees. The pattern produced by the SGLM tended to overestimate abundance at distances >60 m, probably due to saplings frequently found far from adults. Such scattering of saplings likely results from secondary dispersal of *E.grandiflora* seeds by rodents, as observed by Forget (1992).

Among population variables, distance to the nearest adult was the most informative variable. Effects of other population variables were generally low. However conspecific trees had a strong effect for *E. falcata*. Figure 5.c shows negative-density dependence at the cell scale <20 m. Such effect is consistent with local dispersal and aggregation of the species along bottomlands.

4.2 *Relevance and limits of the data model*

The *data model* defined how the sampling design characterized sapling pattern and how ecological variables characterized local ecological conditions. The use of quadrat-sampling imposed a minimal distance between observations and because of clumping, abundance data were positively autocorrelated. Yet, autocorrelation also appears with individual based sampling and is actually inherent to ecological data (Legendre, 1993). Space partitioning in sampling cells allowed to define a measure of species local success, through sapling abundance, and to relate it with ecological variables. Coupled with regular sampling, this method introduced zero inflation in the data. The two features could be accounted for in statistical models.

In studies of tropical species, descriptions of the environment often derive from categorical variables, for instance classifying forest structure as gap and non-gap. The environment was characterized here through continuous descriptors of ecological processes such as disturbance (Molino and Sabatier, 2001). Such a method allowed to characterize species responses along gradients. Inclusion of disturbance as an explicative process allowed to address a common issue in static modelling, that is a likely disequilibrium between observed patterns and current environmental conditions (Guisan and Zimmermann, 2000; Austin, 2002).

Dispersal around potential mother-trees was accounted for in our *data model* through distance to nearest adult (**dna**). Measures of early life-stages dispersal with **dna** underestimate dispersal distances given adult spatial patterns (Nathan and Muller-Landau, 2000). More complex dispersal models could be considered for instance by taking into account the contribution of several potential seeders (Ribbens *et al.*, 1994). Other population variables accounting

for interactions with conspecific trees were based on the same 20 m-radius subplots used for the total stand. However, tree may interact at different distances with conspecific trees or with trees of other species (heteromyopia) as suggested by Murrell and Law (2003). Further work could consider different distances of inter- and intraspecific interactions.

4.3 Autocorrelation and zero-inflation in statistical models

The spatial random effect for autocorrelation enhanced models performance, a consistent result with the comparative study of Keitt *et al.* (2002). Mixed models reproduced sapling pattern through a simple description of dependence between cells at local scale. SGLM with Poisson distribution were better for *O.asbeckii*, and *E.grandiflora*, in spite of high zero-inflation for *E.grandiflora*. Thus, the spatial random effect accounted for both autocorrelation and zero-inflation. ZIP and SZIMP comparison also revealed that variables with strong effects in the simple ZIP case could have weaker effects when autocorrelation was considered. Hence, refinement of the statistical hypotheses lead to more robust conclusions regarding the ecological model (Austin, 2002).

Apart from accounting for zero-inflation in abundance data (Welsh *et al.*, 1996; Ridout *et al.*, 1998), a major interest of ZIP models stands in the shape of the response curves, a critical issue in modelling studies (Oksanen and Minchin, 2002). Indeed, the response to a specific variable is not *a priori* fixed as in a GLM. Different shapes are likely to approach linear or more complex species responses to a particular ecological gradient (Guisan and Zimmermann, 2000).

The random spatial process also introduced flexibility and accounted for unobserved local effects at the scale of sampling cells. This was evidenced for *Q.rosea* for which standing deaths (M_{sDL}) influenced sapling abundance positively as shown in both ZIP and SZIMP models. Still, the deterministic relation failed in the ZIP case: extreme values of the ecological variable determined high adjusted sapling abundance in some cells where saplings were absent. Unobserved local effects may be responsible for such departure which was not the case in the SZIMP model for *Q.rosea*.

The sets of selected covariables, \mathbf{P} , (\mathbf{B}, \mathbf{M}), were considered as *a priori* given for each species and we did not address the issue of variable selection. Powerful but time-consuming Bayesian methods exist and are still being developed to deal with variable selection issues (e.g. Reversible Jump, Green (1995)). Selection classically aims to minimize covariance among explicative variables and identify the most relevant among an initial set. Although the

studied sets of variables may not be optimal for some species, the most informative variables were retained and only weakly informative variables among those tested may miss.

5 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 seems 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. The studied sapling pattern showed evidence of niche partitioning among species and also revealed dispersal limitation. Following the modelling framework proposed by Austin (2002), this work addressed the relevance and limits of three hypotheses levels and possible interconnections between these levels. Statistical hypotheses were based on features of the studied patterns, while explicative ecological data were mostly constrained by feasibility and availability. In any ecological study, ecological implications are conditioned by such constraints. The interest of the data model is thus clear when proxies infer the intensity of unobserved processes. More realistic treatments of spatial patterns would certainly refine our understanding of disturbance effects on tropical forests communities.

6 Acknowledgements

We thank Bruno Ferry (ENGREF, Nancy) for kindly providing data on soil hydromorphy, and Lilian Blanc, Jean-Gaël Jourget, Pascal Pétronelli (CIRAD, Kourou, French Guiana) and the field workers of Paracou who participated in data collection.

References

- Austin, M. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157(2-3)**, 101–118.
- Banerjee, S., Carlin, B. and Glefand, A. (2003) *Hierarchical modeling and analysis for spatial data*. Monographs on Statistics and Applied Probability 101. Chapman & Hall/CRC.

- Bar-Hen, A. (2002) Generalized principal component analysis of continuous and discrete variables. *Interstat*, **8(6)**, 1–26.
- Barthes, B. (1991) The effect of soil type on the distribution of two species of *Eperua* (Caesalpinaceae) in French Guiana influence des caracteres pedologiques sur la repartition spatiale de deux especes du genre *Eperua* (Caesalpinaceae) en foret guyanaise. *Revue d'Ecologie, la Terre et la Vie*, **46(4)**, 303–320.
- Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B*, **36**, 192–236.
- Bonjour, I. (1996) *Variabilité de la qualité des bois d'une essence guyanaise : l'Angélique (Dicorynia guianensis Amsh.)*. Ph.D. thesis, ENGREF.
- Clark, D. and Clark, D. (1999) Assessing the growth of tropical rain forest trees: issues for forest modeling and management. *Ecological Applications*, **9(3)**, 981–997.
- Clark, J. (2005) Why environmental scientists are becoming bayesians. *Ecology Letters*, **8**, 2–14.
- Clark, J., Beckage, B., Camill, P., Cleveland, B., Hillerislambers, J., Lichter, J., McLachlan, J., Mohan, J. and Wyckoff, P. (1999) Interpreting recruitment limitation in forests. *American Journal of Botany*, **86(1)**, 1–16.
- Clark, J. and Ji, Y. (1995) Fecundity and dispersal in plant populations: implications for structure and diversity. *The American Naturalist*, **146**, 72–111.
- 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.
- Dalling, J. and Wirth, R. (1998) Dispersal of *Miconia argentea* seeds by the leaf-cutting ant *Atta colombica*. *Journal of Tropical Ecology*, **14(5)**, 705–710.
- Doligez, A. (1996) *Evolution de la diversité génétique intra-population et de sa structure : Etude d'un modèle de simulation spatialisé en vue de la gestion des ressources génétiques forestières tropicales. Volume I : texte*. Ph.D. thesis, Institut National Agronomique Paris-Grignon.
- Favrillon, V. (1995) *Modèle marticiel déterministe en temps discret. Application à l'étude de la dynamique d'un peuplement forestier tropical humide (Guyane française)*. Ph.D. thesis, Université Claude Bernard - Lyon 1.
- Forget, P.-M. (1992) Regeneration ecology of *Eperua grandiflora* (caesalpinaceae), a large-seeded tree in french guiana. *Biotropica*, **24(2a)**, 146–156.

- Gourlet-Fleury, S., Ferry, B., Molino, J.-F., Petronelli, P. and Schmitt, L. (2004) Experimental plots : key features. In *Ecology and Management of a Neotropical Rainforest. Lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Elsevier, Paris, pp. 3–60.
- Gourlet-Fleury, S. and Houllier, F. (2000) Modelling diameter increment in a lowland evergreen rain forest in french guiana. *Forest Ecology and Management*, **131**, 269–289.
- Green, P. J. (1995) Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, **82**, 711–732.
- Grubb, P. (1977) The maintenance of species richness in plant communities: the importance of the regeneration niche. *Biological Reviews*, **52**, 107–145.
- Guisan, A. and Zimmermann, N. E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135(2/3)**, 147–186.
- Harms, K., Condit, R., Hubbell, S. and Foster, R. (2001) Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *Journal of Ecology*, **89**, 947–959.
- He, F., Legendre, P. and LaFrankie, J. V. (1997) Distribution patterns of tree species in a Malaysian tropical rain forest. *Journal of Vegetation Science*, **8(1)**, 105–114.
- Howe, H. (1989) Scatter- and clump-dispersal and seedling demography: hypothesis and implications. *Oecologia*, **79**, 417–426.
- Keitt, T., Bornstad, O., Dixon, P. and Citron-Pousty, S. (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, **25**, 616–625.
- Lambert, D. (1992) Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.
- Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74(6)**, 1659–1673.
- Loubry, D. (1993) Les paradoxes de l'angélique (*dicorynia guianensis amshoff*) : dissémination et parasitisme des graines avant dispersion chez un arbre anémochore de forêt guyanaise. *Rev. Ecol. (Terre Vie)*, **48**, 353–363.
- McCullagh, P. and Nelder, J. (1989) *Generalized Linear Models*. Monographs on Statistics and Applied Probability 37. Chapman & Hall edition.
- Molino, J. F. and Sabatier, D. (2001) Tree diversity in tropical rain forests: A validation of the intermediate disturbance hypothesis. *Science*, **294(5547)**, 1702–1704.
- Murrell, D. J. and Law, R. (2003) Heteromyopia and the spatial coexistence of similar competitors. *Ecology Letters*, **6(1)**, 48–59.
- Nathan, R. and Muller-Landau, H. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology and*

- Evolution*, **15(7)**, 278–285.
- Nicotra, A. B., Chazdon, R. L. and Iriarte, S. V. B. (1999) Spatial heterogeneity of light and woody seedling regeneration in tropical forests. *Ecology*, **80(6)**, 1908–1926.
- Oksanen, J. and Minchin, P. (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, **157(2-3)**, 119–129.
- Pélissier, R. and Goreaud, F. (2001) A practical approach to studying the spatial structure in simple cases of heterogeneous vegetation stands. *Journal of Vegetation Science*, **12**, in press.
- Plotkin, J., Chave, J. and Ashton, P. (2002) Cluster analysis of spatial patterns in Malaysian tree species. *The American Naturalist*, **160(5)**, 629–644.
- R Development Core Team (2004) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rankin-de Mérona, J. and Montpied, P. (2004) Natural regeneration of selected tropical rain forest tree species of french guiana: established seedlings and saplings during the period 1986-1995. In *Ecology and Management of a Neotropical Rainforest - Lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Elsevier, Paris, pp. 194–218.
- Ribbens, E., Silander, J. and Pacala, S. (1994) Seedling recruitment in forests: calibrating models to predict patterns of seedling dispersion. *Ecology*, **75(6)**, 1794–1806.
- Ricklefs, R. (1977) Environmental heterogeneity and plant species: a hypothesis. *American Naturalist*, **111**, 377–381.
- Ridout, M., Demetrio, C. and Hinde, J. (1998) Models for count data with many zeros. In *International Biometric Conference*. Cape Town.
- Rushton, S., Ormerod, S. and Kerby, G. (2004) New paradigms for modelling species distribution? *Journal of Applied Ecology*, **41**, 193–200.
- Russo, S. and Augspurger, C. (2004) Aggregated seed dispersal by spider monkeys limits recruitment to clumped patterns in *Virola calophylla*. *Ecology Letters*, **7(11)**, 1058–1067.
- Sabatier, D. (1983) *Fructification et dissémination en forêt guyanaise - L'exemple de quelques espèces ligneuses*. Doctorat de 3ème cycle, Université des Sciences et Techniques du Languedoc.
- Sheil, D. (1999) Tropical forest diversity, environmental change and species augmentation: After the intermediate disturbance hypothesis. *Journal of Vegetation Science*, **10**, 851–860.
- Shugart, H. (1984) *A Theory of Forest Dynamics. The Ecological Implications of Forest Succession Models*. Springer Verlag, New York.

- Spiegelhalter, D. (2004) *WinBUGS, Bayesian Inference Using Gibbs Sampling*. Imperial College and Medical Research Council, London, UK. ISBN 3-900051-07-0.
- Spiegelhalter, D., Best, N., Carlin, B. and Van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **6**, 583–639.
- Svenning, J.-C. (2001) Environmental heterogeneity, recruitment limitation and the mesoscale distribution in a tropical montane rain forest (maquipucuna, ecuador). *Journal of Tropical Ecology*, **17**, 97–113.
- Ulft, L. v. (2004) *Regeneration in Natural and Logged Tropical Rain Forest. Modelling seed dispersal and regeneration of tropical trees in Guyana*, volume 12 of *Tropenbos-Guyana Series*. Tropenbos International, Georgetown.
- Valencia, R., Foster, R., Villa, G., Condit, R., Svenning, J., Hernandez, C., Romoleroux, K., Losos, E., Magard, E. and Balslev, H. (2004) Tree species distributions and local habitat variation in the amazon: large forest plot in eastern ecuador. *Journal of Ecology Oxford*, **92(2)**, 214–229.
- Wang, B. and Smith, T. (2002) Closing the seed dispersal loop. *Trends in Ecology and Evolution*, **17(8)**, 379–385.
- Welsh, A., Cunningham, R., Donnelly, C. and Lindenmayer, D. (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, **88(1-3)**, 297–308.
- Wikle, C. (2003) Hierarchical bayesian models for predicting the spread of ecological processes. *Ecology*, **84**, 1382–1394.
- Wikle, C. and Anderson, C. (2003) Climatological analysis of tornado report counts using a hierarchical bayesian spatio-temporal model. *Journal of Geophysical Research - Atmospheres*, **108**.
- Zorn, C. (1996) Evaluating zero-inflated and hurdle models specifications. *Midwest Political Science Association, April 18-20*.

Table 1

Ecological variables derived from a DEM (Digitalized Elevation Model) of Paracou, GIS layers of seasonal streams and bottomlands, skid trails and logging damage, or from census data of trees ≥ 10 cm DBH (*units* in brackets). For quantitative variables, the period indicates calculus years: 1986-1988 (*logging*) or 1989-2002 (*recovery*). Static variables were calculated in 2002. **diam1** and **diam2** derived from a Canonical Analysis of the numbers of trees in 11 DBH classes from 10-15 cm to 55-60 cm and >60 cm DBH (16 and 14 % of total inertia explained). **diam1** (resp. **diam2**) separated subplots with well represented *low-size* classes (resp. *middle-size* classes, positive scores) from subplots with over-represented *large-size* classes (resp. *extreme-size* classes, negative scores).

Type	Label	Description	Period
<i>Physiography</i>	Ele	Elevation (<i>m</i>)	-
	Slo	Slope ($^{\circ}$)	-
<i>Logging disturbance</i>	WL	1 : outside bottomlands 2 : 20 m buffer along bottomlands, and 10 m along seasonal streams 3 : bottomlands (watertable < 1 m depth during dry season)	-
	ST	1 : >10 m from a skid trail 2 : 10 m buffer around skid trails 3 : skid trails	-
	LD	1 : > 10 m from logging damage 2 : 10 m buffer around logging damage 3 : logging damage	-
<i>Post-logging dynamics</i>	M _{tfL}	Basal area lost in treefalls (m^2)	1986-1988
	M _{sdL}	Basal area lost in standing deaths (m^2)	
	M _{tfR}	Basal area lost in treefalls (m^2)	1989-2002
A _{tf}	Mean age of treefalls (<i>yr</i>)		
SD _{tfR}	Standard deviation of treefalls ages (<i>yr</i>)		
M _{sdR}	Basal area lost in standing deaths (m^2)		
<i>Structure</i>	Recru	Basal area of recruited individuals (m^2)	2002
	dG	Change in basal area (m^2)	
	G _{pio}	Basal area of pioneer taxa (m^2)	
	diam1	Axis1 of CA on diameter distributions	
<i>Population variables</i>	diam2	Axis2 of CA on diameter distributions	1989-2002
	G _{tot}	Total basal area (m^2)	
	dna	Distance to nearest adult (<i>m</i>)	2002
<i>Population variables</i>	dG _{con}	Loss of basal area from conspecific trees death (m^2)	1989-2002
	G _{con}	Basal area of conspecific trees ≥ 10 cm dbh (m^2)	2002

Table 2

Summary statistics of calibrated models. p : number of regression parameters, p_D : number of effective parameters, DIC : Deviance Information Criterion, r_P , r_S : Pearson and Spearman correlation coefficients between observations and fitted values. **Bold**: lower DIC values by species. *Oa*: *O.asbeckii*, *Ef*: *E.falcata*, *Eg*: *E.grandiflora*, *Dg*: *D.guianensis*, *Qr*: *Q.rosea*, *Tm*: *T.melinonii*. SGLM: Spatial Generalized Linear Mixed model, ZIP: Zero Inflated Poisson model, SZIMP: Spatial Zero Inflated Mixed model.

Model	Species					
	<i>Oa</i>	<i>Ef</i>	<i>Eg</i>	<i>Dg</i>	<i>Qr</i>	<i>Tm</i>
SGLM						
p	20	15	18	14	23	19
p_D	1246.0	770.8	908.3	751.6	708.5	791.7
DIC	5907.0	2742.8	3257.6	2646.6	2573.8	2493.4
r_P	0.92	0.94	0.92	0.92	0.98	0.95
r_S	0.76	0.61	0.64	0.59	0.69	0.55
ZIP						
p	31	25	24	20	32	30
p_D	3547.1	706.7	2533.1	2436.4	781.0	1247.5
DIC	8142.6	3067.8	5019.0	4310.0	2816.9	2744.3
r_P	0.48	0.57	0.48	0.23	0.22	0.24
r_S	0.48	0.53	0.45	0.31	0.23	0.26
SZIMP						
p	33	27	26	22	34	32
p_D	1266.7	621.2	917.8	755.6	694.7	775.1
DIC	5907.4	2499.3	3263.1	2629.0	2556.2	2462.7
r_P	0.92	0.93	0.90	0.92	0.96	0.95
r_S	0.76	0.61	0.63	0.59	0.68	0.55

Appendices

A Site map

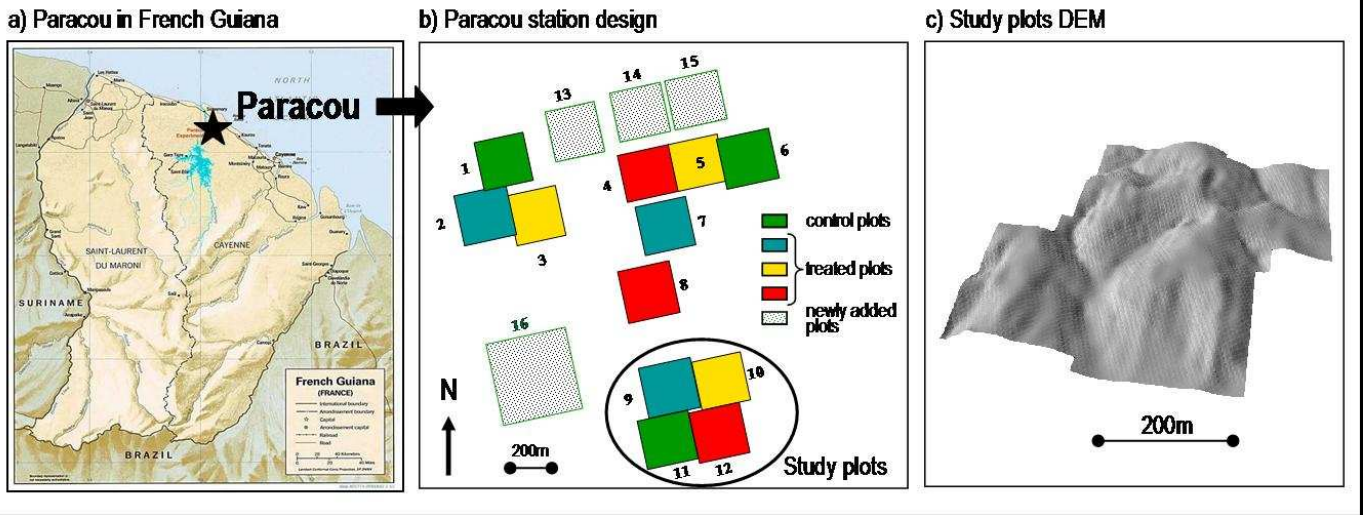


Fig. A.1. Map of the study site

B Abundance maps

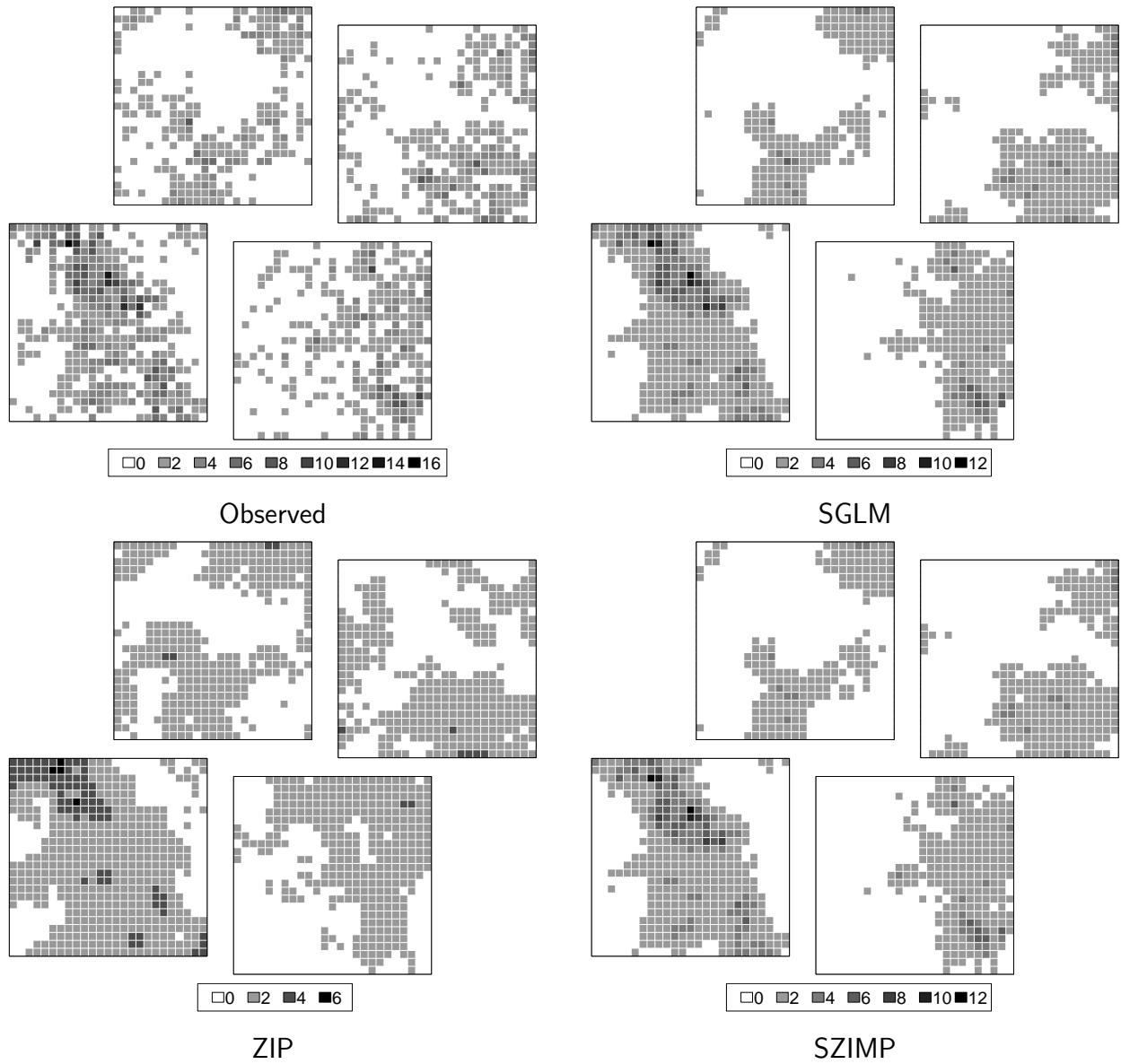


Fig. B.1. Observed (*top left*) and modelled sapling abundance of *O. asbeckii* in sampling cells in the four studied PSP.

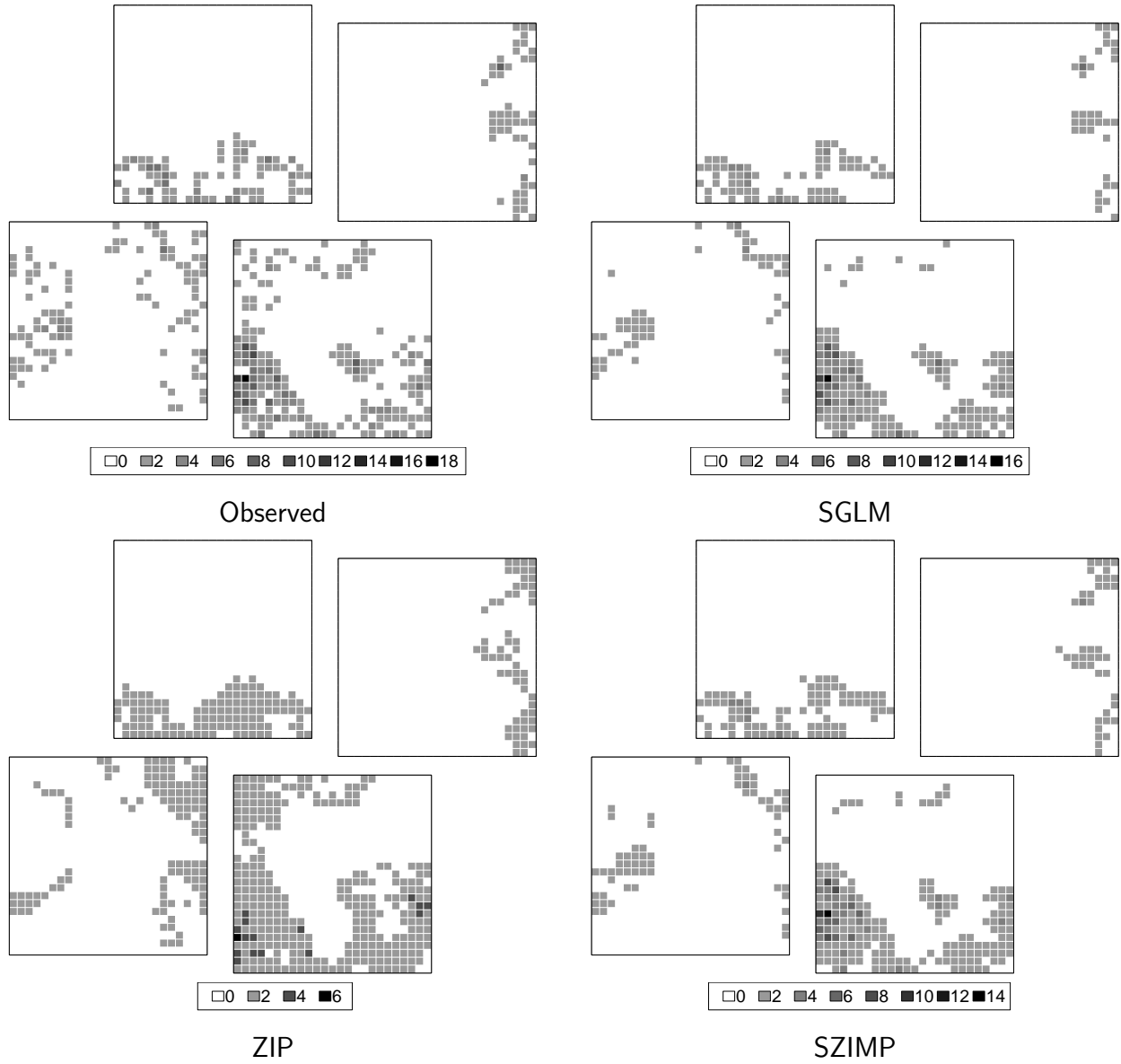


Fig. B.2. Observed (*top left*) and modelled sapling abundance of *E.falcata* in sampling cells in the four studied PSP.

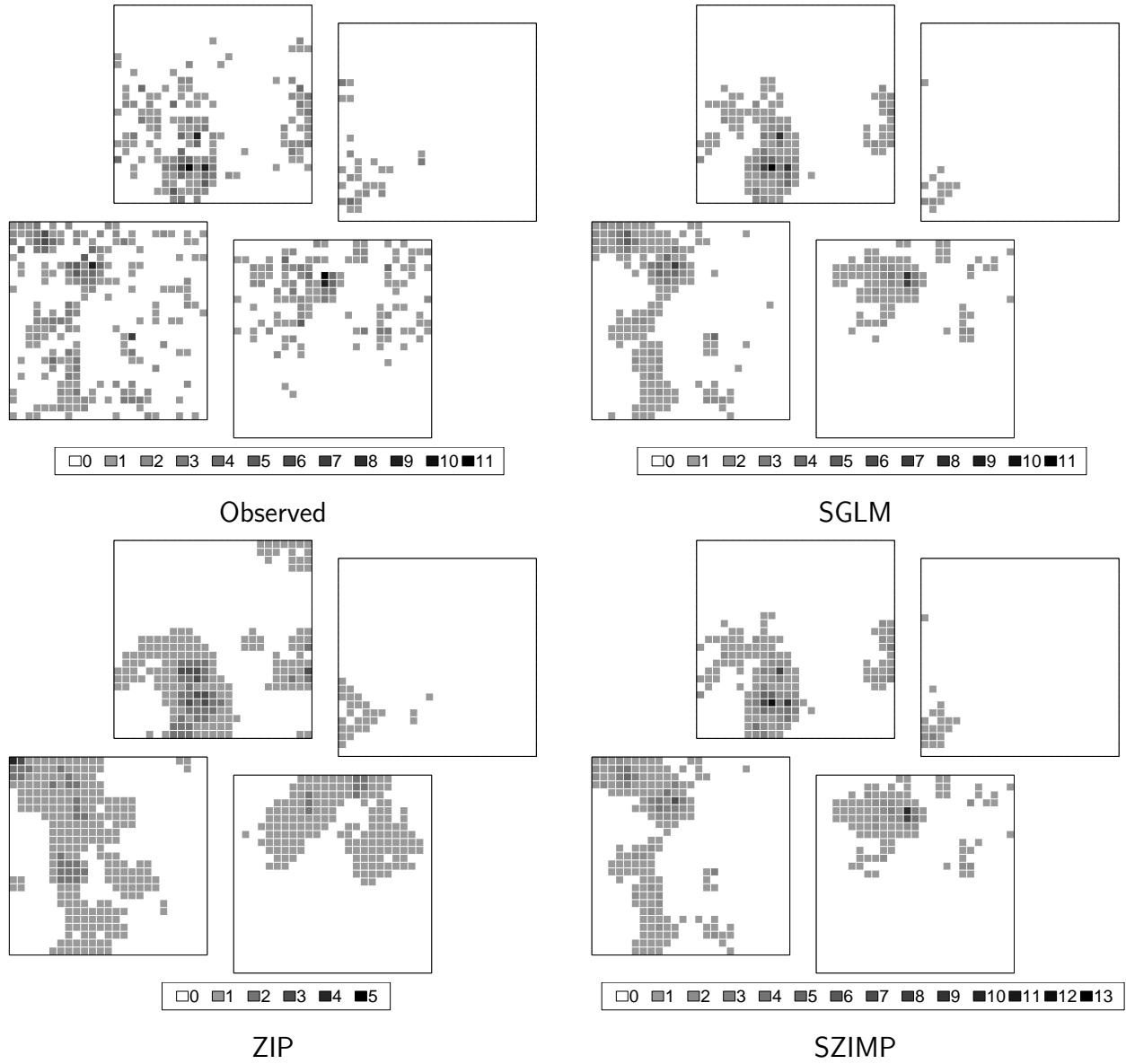


Fig. B.3. Observed (*top left*) and modelled sapling abundance of *E. grandiflora* in sampling cells in the four studied PSP.

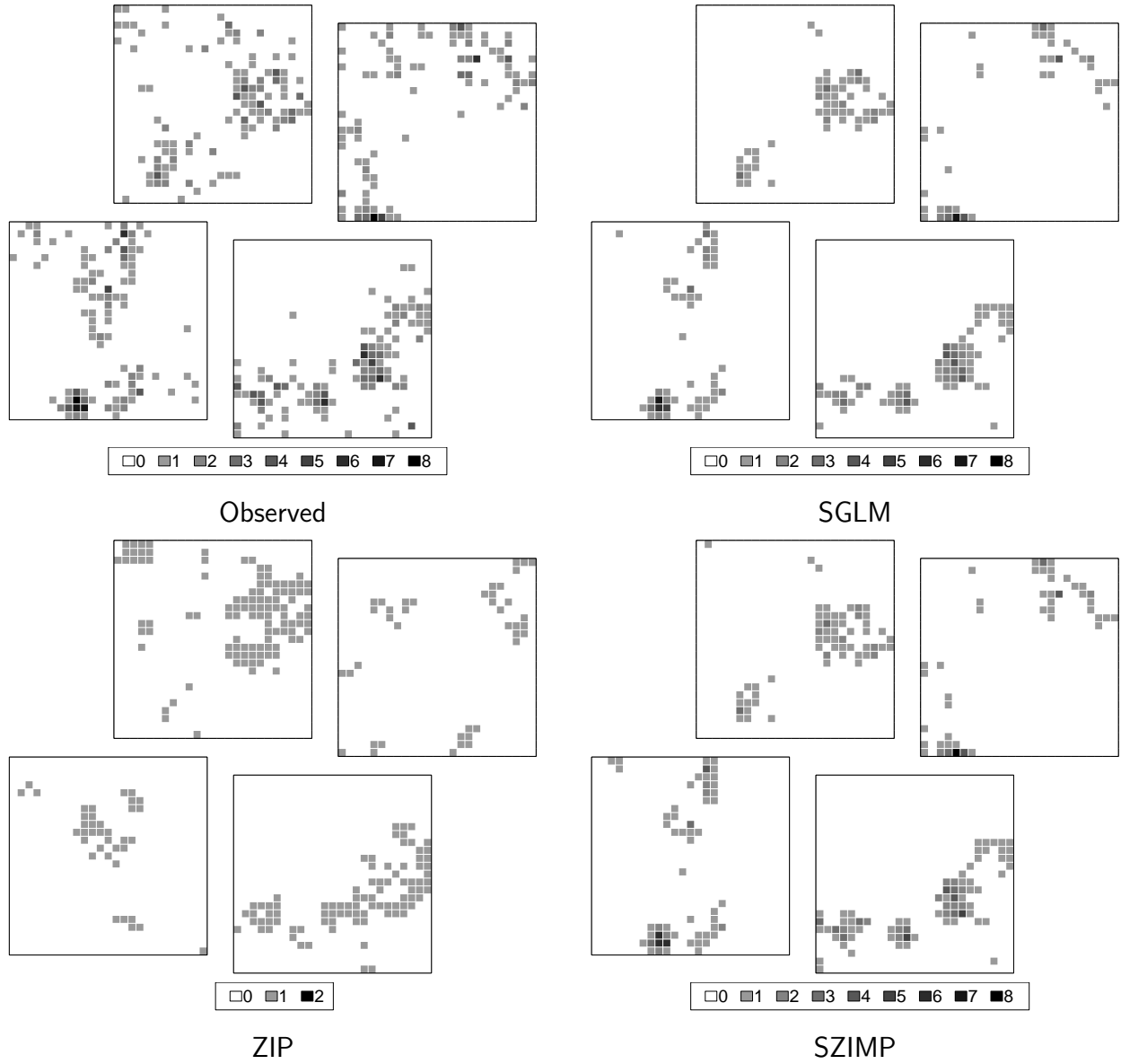


Fig. B.4. Observed (*top left*) and modelled sapling abundance of *D. guianensis* in sampling cells in the four studied PSP.

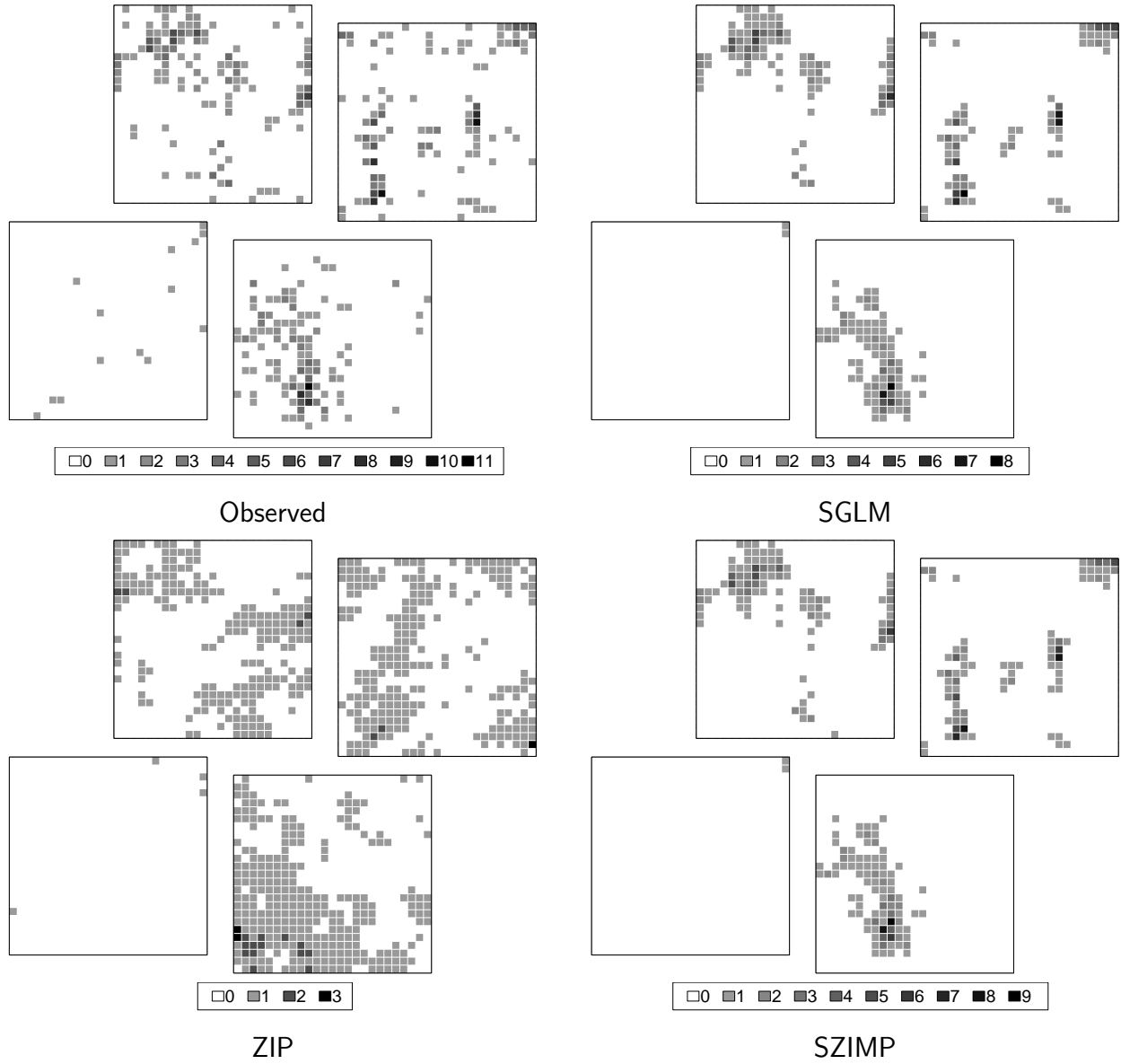


Fig. B.6. Observed (*top left*) and modelled sapling abundance of *T.melinonii* in sampling cells in the four studied PSP.