



# Environmental control of natural gap size distribution in tropical forests

Youven Goulamoussène<sup>1</sup>, Caroline Bedeau<sup>2</sup>, Laurent Descroix<sup>2</sup>, Laurent Linguet<sup>1</sup>, and Bruno Hérault<sup>3</sup>

<sup>1</sup>Université de Guyane – UMR Espace-Dev, BP 792, 97337 Cayenne, French Guiana

<sup>2</sup>Office National des Forêts (ONF), Département RD, Cayenne, French Guiana

<sup>3</sup>Cirad, UMR EcoFoG (AgroParisTech, CNRS, Inra, Univ Antilles, Univ Guyane), Kourou, French Guiana

*Correspondence to:* Youven Goulamoussène (youven.goulamoussene@ecofog.gf)  
and Bruno Hérault (bruno.herault@cirad.fr)

Received: 28 July 2016 – Published in Biogeosciences Discuss.: 8 August 2016

Revised: 21 December 2016 – Accepted: 3 January 2017 – Published: 24 January 2017

**Abstract.** Natural disturbances are the dominant form of forest regeneration and dynamics in unmanaged tropical forests. Monitoring the size distribution of treefall gaps is important to better understand and predict the carbon budget in response to land use and other global changes. In this study, we model the size frequency distribution of natural canopy gaps with a discrete power law distribution. We use a Bayesian framework to introduce and test, using Monte Carlo Markov chain and Kuo–Mallick algorithms, the effect of local physical environment on gap size distribution. We apply our methodological framework to an original light detecting and ranging dataset in which natural forest gaps were delineated over 30 000 ha of unmanaged forest. We highlight strong links between gap size distribution and environment, primarily hydrological conditions and topography, with large gaps being more frequent on floodplains and in wind-exposed areas. In the future, we plan to apply our methodological framework on a larger scale using satellite data. Additionally, although gap size distribution variation is clearly under environmental control, variation in gap size distribution in time should be tested against climate variability.

with younger trees (Whitmore, 1989). The first, and perhaps most important, effect of gap occurrence is an immediate increase in light intensity (Hubbell et al., 1999), allowing sunlight to penetrate the understory. This phenomenon has been widely studied because the opening of gaps contributes to the establishment and growth of light-demanding trees (Denslow et al., 1998), thus contributing to the maintenance of biodiversity. Another effect of canopy gaps is the local modification of the forest nutrient balance (Rüger et al., 2009). When canopy gaps are created, large amounts of dead leaves and wood will be decomposed and mineralized so that the availability of soil nutrients for neighboring trees will increase (Brokaw, 1985). These nutrient patches are also linked to small-scale spatial variations in forest carbon balance, as shown by (Feeley et al., 2007). The relationship between gap formation and the population dynamics of trees or lianas is also quite well understood, with increased liana basal area (Schnitzer et al., 2014) and low-wood-density pioneer species that recruit exclusively in newly formed gaps (Molino and Sabatier, 2001).

Many studies have investigated the effect of treefall gaps on biodiversity, particularly animal communities (Bicknell et al., 2014; Puerta-Piñero et al., 2013), on the carbon cycles, and on forest dynamics. Some authors use field data to study natural gap dynamics, usually at plot scale (Hubbell et al., 1999). As these studies are quite limited in spatial extent (< 50 ha) and because gap formation is largely unpredictable (Hubbell et al., 1999; Lloyd et al., 2009), optical satellite imagery has been widely promoted and proven adequate for monitoring forest gaps over space

## 1 Introduction

Natural disturbances caused by forest gaps play an important role in tropical rainforest dynamics. Canopy gaps caused by the death of one or more trees are the dominant form of forest regeneration because the creation of canopy openings continuously reshapes forest structure as gaps are filled

and time (Frolking et al., 2009). At high resolution (< 10 m), IKONOS satellite images may be well suited for evaluating gap dynamics (Espírito-Santo et al., 2014). In French Guiana, the SPOT-4 satellite (20 m spatial resolution) has successfully detected canopy gaps (Colson et al., 2006) using a combination of several spectral bands, such as near and short-wave infrared. However, topographical variation, gap shape, and shade may influence and bias gap detection with optical products. Moreover, persistent cloud cover, which is common in many tropical forests, limits their utility.

Airborne light detecting and ranging (lidar) platforms therefore offer a solution to this problem. Recent developments in lidar have significantly advanced our ability to derive accurate measurements of canopy forest structure, to detect gaps, and to assess the effect of spatial and temporal variation on carbon balance (Asner and Mascaro, 2014). (Kellner and Asner, 2009) used remote lidar sensing to quantify canopy height and gap size distributions in five tropical rain forest landscapes in Costa Rica and Hawaii. They showed that canopy gaps can be observed with the help of lidar-derived digital canopy models (DCMs) and that gap size frequency distribution (GSFD) can be fit with a power law distribution, suggesting a surprising similarity in canopy gap size frequency distributions on diverse soil types with different geologic substrate ages. (Asner et al., 2013) also used lidar data to analyze whether gap size frequency distribution is modified by topographic and geologic characteristics and again showed that canopy gap size distribution is largely invariant between forests on erosional terra firme and depositional floodplain substrates in the Peruvian Amazon basin. Finally, using airborne lidar, Lobo and Dalling (2014) have recently explored the effect of forest age, topography, and soil type on canopy disturbance patterns across central Panama. For the first time, they highlighted significant effects of slope and of forest age, with a higher frequency of large gaps associated with old-growth forests and gentle slopes.

In the present study, we use a DCM derived from airborne lidar across a 30 000 ha tropical forest landscape in the Régina forest in French Guiana. This approach provides high-resolution maps of canopy gaps and helps us to understand the environmental determinism of gap occurrence in tropical forests. Our specific aims were therefore

- to define canopy gaps from canopy height data using a probabilistic approach;
- to model gap size distribution by inferring a likelihood-explicit discrete power law distribution in a Bayesian framework; and
- to introduce the environment into the scaling parameter of the power law distribution and test its predictive ability.

## 2 Materials and methods

The study site is located in the Régina forest (4° N, 52° W), where the most common soils are ferralitic. The site is located on slightly contrasting plateau-type reliefs that are rarely higher than 150 m on average. The forest is typical of French Guianese rainforests. Dominant plant families in the Régina forest include *Burseraceae*, *Mimosoideae*, and *Caesalpinoideae*. The site receives 3806 mm of precipitation per year, with a long dry season from mid-August to mid-November and a short dry season in March (Wagner et al., 2011).

### 2.1 Data source

#### 2.1.1 Lidar data

Lidar data were acquired by aircraft in 2013 over 30 000 ha of forest by a private contractor, Altoa (<http://www.altoa.fr/>), using a Riegl LMS-Q560 laser. This system was composed of a scanning laser altimeter with a rotating mirror, a GPS receiver (coupled to a second GPS receiver on the ground), and an inertial measurement unit to record the pitch, roll, and heading of the aircraft. The laser wavelength was near-infrared (from about 800 to 2500 nm). Flights were conducted at 500 m above ground level with a ground speed of 180 km h<sup>-1</sup>, and each flight derived two acquisitions. The lidar was operated with a scanning angle of 60° and a 200 kHz pulse repetition frequency. The laser recorded the last reflected pulse with a precision better than 0.10 m, with a density of 5 pulses m<sup>-2</sup>.

The DCM was derived from the raw scatter plot consisting of the pooled dataset from the two acquisitions. Raw data points were first processed to extract ground points using the TerraScan (TerraSolid, Helsinki) ground routine, which classifies ground points by iteratively building a triangulated surface model. Ground points typically made up less than 1 % of the total number of the return pulses. The DCM has a resolution of 1 m. In order to avoid delineating “false” gaps due to river beds, we remove areas very close to natural rivers with a 20 m buffer applied to all shorelines. Additionally, a 25 m buffer was applied to exclude anthropogenic tracks.

#### 2.1.2 Environmental data

We use six environmental variables to synthesize the observed environmental gradients. All variables were computed from a lidar digital terrain model (DTM) with 5 m<sup>2</sup> cells.

#### Slope

The slope was derived from the lidar DTM. Slope was computed at a grid cell as the maximum rate of change in elevation from that cell to its eight neighboring cells over the distance between them.

## Topographic exposure

We use the TOPographic EXposure (TOPEX) index to measure topographic exposure to wind (Chapman, 2000). TOPEX is a variable that represents the degree of shelter assigned to a location. It was derived from quantitative assessment of horizon inclination. The values of this index are closely correlated with the wind-shape index (Mikita and Klimánek, 2012). Exposure is calculated based on the height and distance of the surrounding horizon, which are combined to obtain the inflection angle. We use this angle to quantify topographic exposure (pixel resolution 5 m × 5 m). When a large topographic feature, e.g., a mountain, is far off in the distance, the inflection angle is low. When the same mountain is closer, the inflection angle is higher. Therefore, a higher inflection angle is equal to lower exposure or higher sheltering (Mikita and Klimánek, 2012).

## Drained area

Drained area (DA) measures the surface of the hydraulic basin that flows through a cell. A low value indicates that a cell is located at the border between two basins, whereas high values indicate cells located downstream.

## Hydraulic altitude

The hydraulic altitude (HA) of each cell, its altitude above the closest stream of its hydraulic basin, was computed from the third-order hydraulic system. Low values, including 0, indicate that the forest plot is potentially temporarily flooded, whereas high values indicate that it is located on a hilltop.

## Terrain ruggedness index

The terrain ruggedness index (TRI) captures the difference between flat and mountainous landscapes. TRI was calculated using SAGA GIS (SAGA, 2013) as the sum of the altitude change between a pixel and its eight neighboring pixels (Riley, 1999).

## The height above the nearest drainage

The height above the nearest drainage (HAND) model normalizes topography with respect to the drainage network by applying two procedures to the DTM. The initial basis for the HAND model came from the definition of a drainage channel: perennial streamflow occurs at the surface, where the soil substrate is permanently saturated. It follows that the terrain at and around a flowing stream must be permanently saturated, independently of the height above sea level at which the channel occurs. Streamflow indicates the localized occurrence of homogeneously saturated soils across the landscape. The second basis for the HAND model came from the distinctive physical features of water circulation. Land flows proceed from the land to the sea in two phases: in restrained

flows at the hillslope surface and subsurface, and in freer flows (or discharge) along defined natural channels (Nobre et al., 2011).

## 2.2 Forest gap definition

### 2.2.1 Height threshold

To identify discrete canopy gaps, we had to choose a gap threshold height. Some authors define this threshold at 2 m (Brokaw, 1982). (Runkle, 1982) defines a gap as the ground area under a canopy opening that extends to the base of the surrounding canopy trees, these usually being considered to be taller than 10 m, with a trunk diameter at breast height (DBH) > 20 cm. However, in practice, defining gap boundaries is a tricky issue, even in the field. Here, we develop a probabilistic method for detecting canopy gaps from lidar data. We used the DCM to model canopy height distribution considering a mixture distribution of two ecological states: the natural variation of canopy height in mature forests, modeled as a normal distribution, and the presence of forest gaps, which lead to a new normal distribution with lower values. We consider that the threshold between the two states is equal to the 0.001th percentile of the height distribution of the canopy; our results appeared robust to the threshold value (see Appendix A). We then define canopy gaps as contiguous pixels (in contact by edges or by vertices) at which the height is less than or equal to the height threshold.

### 2.2.2 Minimum gap size

In our study, we define the minimum area of a gap as  $x_{\min}$ . We model the gap size frequency distribution with a power law distribution. We use the Pareto distribution in a discrete power law probability density function (Virkar and Clauset, 2014). These distributions have a negative slope and their size frequencies are plotted on a logarithmic scale, allowing us to infer the scaling parameter  $\lambda$ . A value close to 1 means a large number of large gaps. In other words, in forests dominated by small canopy openings, values of  $\lambda$  are larger, whereas smaller values of  $\lambda$  indicate an increased frequency of large gap events (Fisher et al., 2008). In a discrete power law with parameter  $\lambda$ , the probability for gap size  $x$  is given by

$$p(x) = \frac{x^{-\lambda}}{\zeta(x_{\min}, \lambda)}, \quad (1)$$

where  $x_{\min}$  is the lower truncation point and  $\lambda$  is the scaling parameter.

The statistical analyses were performed in R (R Core Team, 2013), making use of `powerlaw` (Clauset et al., 2009) and `VGAM` (Yee, 2010) packages.

We use a Kolmogorov–Smirnov (KS) distance criterion order to determine the error between the observed distribution and the Pareto distribution. KS is defined as the max-

imum distance between the cumulative distribution functions (CDFs) of the data and the fitted function (Virkar and Clauset, 2014). We retain, for the remainder of this study, a minimum gap size area  $x_{\min} = 104 \text{ m}^2$ , which minimized the KS distance in our dataset.

### 2.3 Modeling gap size distribution

Having set the height threshold and minimum gap size, the GSFD is modeled with a discrete Pareto distribution frequency.

#### 2.3.1 Model inference

We use a Bayesian framework to estimate model parameters. Here, the value of a parameter is estimated by its posterior distribution, which by definition is proportional to the product of the likelihood of the model and the parameter prior distribution. The prior distribution is based on prior knowledge of the possible values of a parameter. The posterior densities of the different parameters were estimated using a Monte Carlo Markov chain algorithm (MCMC).

#### 2.3.2 Metropolis–Hastings algorithm

As the model contains many parameters, we built a Metropolis–Hastings (MH) algorithm in which all parameters are updated together. Details on the algorithm are given below:

- $Y = y_1, y_2, \dots, y_n$  is the gap size vector;
- $X = x_{g1}, x_{g2}, \dots, x_{gi}$  is the vector of covariates (environmental variables) for gap  $g$ ; and
- $\theta = \theta_1, \theta_2, \dots, \theta_i$  is the model parameter vector.

The first values of the parameter vector are initialized as  $t = 1, \theta^t \sim \pi_\theta^0$ .

For each step  $t$ , a new parameter value is sampled from the proposition distribution and a new vector of theta candidates is generated.

$$\theta^{\text{cand}} \sim \pi^{\text{prop}} \tag{2}$$

Acceptance or rejection of the new candidate  $\theta^{\text{cand}}$  is determined by computing the likelihood ratio of the two discrete Pareto distributions:

$$\rho(\theta^t, \theta^{\text{cand}}) = \frac{\underbrace{\mathcal{L}(Y|X, \theta^{\text{cand}})}_{\text{likelihood}} \underbrace{\pi_\theta^0(\theta^{\text{cand}})}_{\text{prior}} \underbrace{\pi^{\text{prop}}(\theta^t)}_{\text{proposal}}}{\underbrace{\mathcal{L}(Y|X, \theta^t)}_{\text{likelihood}} \underbrace{\pi_\theta^0(\theta^t)}_{\text{prior}} \underbrace{\pi^{\text{prop}}(\theta^{\text{cand}})}_{\text{proposal}}} \tag{3}$$

The candidate  $\theta^{\text{cand}}$  is accepted or rejected as follows:

$$u \sim \mathcal{U}_{(0,1)}, \theta^{\text{cand}} \begin{cases} \theta^{t+1} & \text{if } u < \rho(\theta^t, \theta^{\text{cand}}), \\ \theta^t & \text{if } u > \rho(\theta^t, \theta^{\text{cand}}). \end{cases} \tag{4}$$

The algorithm is run for 1000 iterations. We use the median of the posterior densities to estimate parameter values, and the distribution of the posterior densities to estimate parameter credibility intervals.

#### 2.3.3 Univariate environmental effects

##### Variable transformation

To improve model inference, parameter significance and interpretation, we first transformed some environmental variables:

$$\text{slope} = \text{sqrt}(\text{slope}) \tag{5}$$

$$\text{HALt} = \log(\text{HA} + 1) \tag{6}$$

$$\text{TOPEX} = |\max(\text{TOPEX}) - (\text{TOPEX})| \tag{7}$$

The environmental variables are then centered and scaled with R function “scale”.

We first consider each environmental covariate independently. These covariates are included one-by-one in the model to constrain the exponent  $\lambda$ . We use the exponential function to constrain  $\lambda$ , because the Riemann zeta function only admits  $\lambda > 1$ .

$$\lambda_{ig} = 1 + \exp(\theta_0 + \theta_i \times \text{var}_{ig}), \tag{8}$$

where  $\lambda_{ig}$  is the  $\lambda$  value dependent on the value of environmental variable  $i$  in gap  $g$ ,  $\theta_0$  is the intercept, and  $\theta_i$  quantifies the effect of covariates var on the gap size distribution  $\text{var}_{ig}$ .

#### 2.3.4 Multivariate model

##### Principal component analysis

We first investigated the collinearity of environmental data through principal component analysis (PCA) of the normalized environmental dataset.

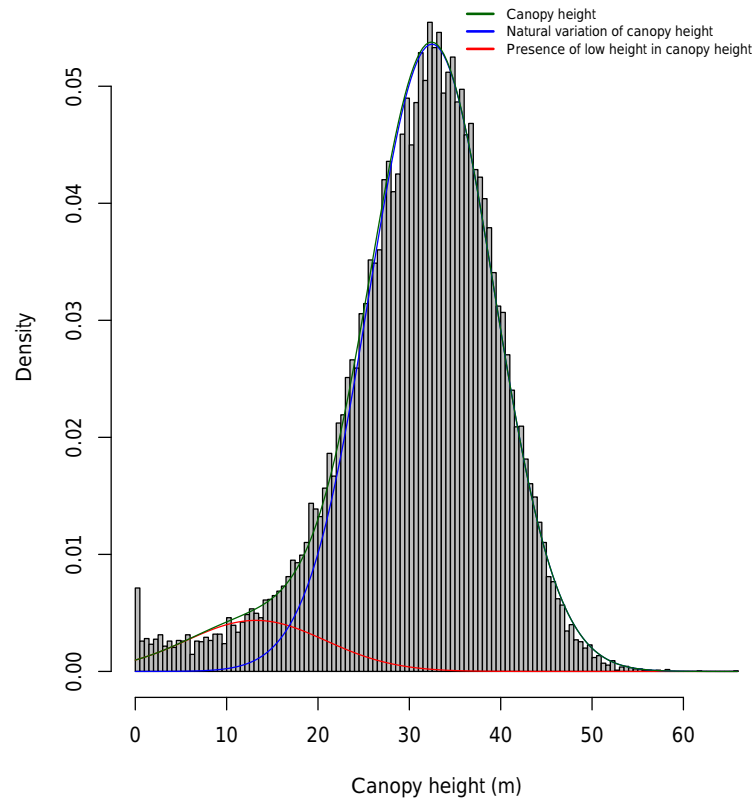
##### Model

To build the final model, we used the results of the univariate model (Table 1) and the PCA (Fig. 2) and set

$$\lambda = 1 + \exp(\theta_0 + \theta_1 \times \text{Slope} + \theta_2 \times \text{TOPEX} + \theta_3 \times \text{HALt} + \theta_4 \times \text{HAND}). \tag{9}$$

##### Variable selection

To select the significant covariates and build the final model, we used the method proposed by (Kuo and Mallick, 1998) (KM). This method consists of associating an indicator with each variable  $\text{var}_i$  and parameter  $\theta_i$ . This indicator can take two values: 1 or 0. If it is set to 1, the variable is included in the model, but if the value is set to 0, it is not. We used



**Figure 1.** Canopy height distribution. Canopy height considered as a mixture distribution of two ecological features. The first (blue curve) is the natural variation in canopy height, modeled as a normal distribution. The second (red curve) is linked to the presence of low heights in the total canopy height distribution, likely to be due to a forest gap. We set the gap threshold to the 0.001th percentile of the blue curve density, i.e., 11 m.

**Table 1.** List of environmental variables, abbreviations, units, and values of the posteriors in univariate models.

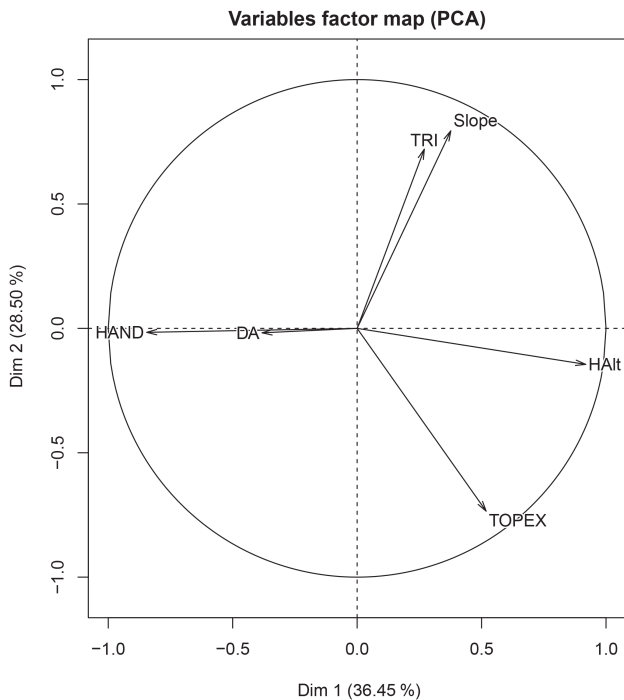
Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95%)
Slope	Slope	°	0.0735	[−0.02; 0.15]
Terrain ruggedness index	TRI	–	0.0718	[0.04; 0.10]
TOPographic EXposure	TOPEX	–	−0.082	[−0.12; −0.05]
Drained area	DA	m <sup>2</sup>	−0.0176	[−0.09; 0.05]
The hydraulic altitude	HA	m	−0.0177	[−0.05; 0.02]
HAND	HAND	–	−0.003	[−0.08; 0.09]

the MH and KM algorithms to estimate the indicators  $I$  and infer their a posteriori distribution in addition to  $\theta$ .

We start the KM algorithm with  $t = 1$ ,  $\theta^t \sim \pi_\theta^0$ ,  $I_j^t \sim \text{Ber}(0.5)$  for  $j = 1, \dots, i$ . For each covariate  $j$  (selected in random order), we use the MH algorithm to update  $\theta_j$ . To update  $I_j$ , we compute the ratio  $\rho$  (Eq. 10) and generate  $I_j^{t+1}$  from a Bernoulli distribution  $\text{Bern}(\rho)$ :

$$\rho = \frac{1}{1 + \frac{\mathcal{L}(Y|\tilde{X}, \theta^t, I_j=0, I_{-j}^t)}{\mathcal{L}(Y|\tilde{X}, \theta^t, I_j=1, I_{-j}^t)}} \quad (10)$$

Model inference and data analysis were conducted with R software (R Core Team, 2012). All maps and geographical information were computed with SAGA (SAGA, 2013) and ArcGIS 10.1.



**Figure 2.** Results of the principal component analysis of the environmental variables.

### 3 Results

#### 3.1 Gap delineation

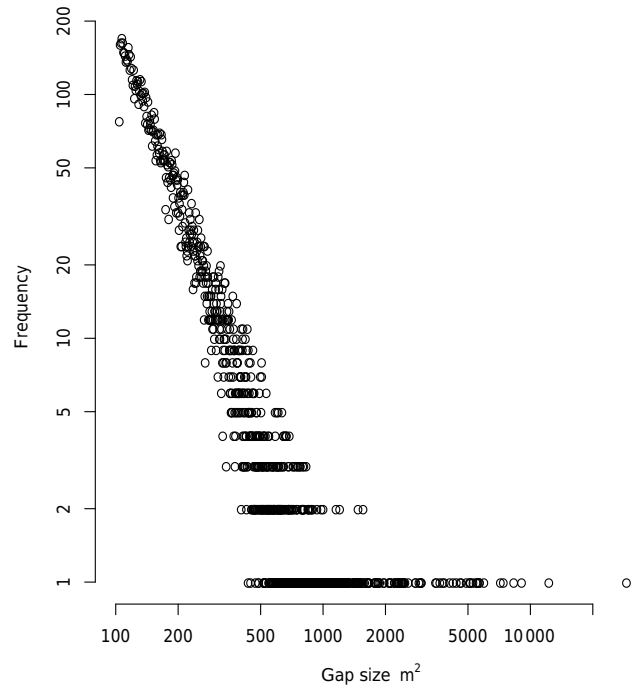
In this study, we used a forest canopy height mixture model to define the maximum height of a given pixel to be included in a forest gap. This probabilistic method produced results that fit the observed canopy height distribution. We retained the 11 m threshold that corresponds to the 0.001th percentile of the canopy height distribution (Fig. 1). Given this height, we retained the surface  $x_{\min} = 104 \text{ m}^2$  that minimized the KS distance between predictions and observations. Here, our gap definition was therefore defined as an area  $> 104 \text{ m}^2$ , in which the lidar measured canopy height is always  $\leq 11 \text{ m}$ .

#### 3.2 Basic statistics

We mapped 12 293 gaps with vegetation  $\leq 11 \text{ m}$  in height. The mean gap size was  $236 \text{ m}^2$  with a minimum gap size of  $104 \text{ m}^2$  and a maximum of  $29\,063 \text{ m}^2$ . The total gap area was about 290 ha, or 1 % of the whole surveyed area. The observed gap size distribution was modeled with a Pareto distribution (Fig. 3), leading to a scaling parameter  $\lambda_{x_{\min}}$  of 2.6.

#### 3.3 Univariate models

All variables had an effect on gap size distribution (Table 1). The scaling coefficient  $\lambda$  is related to the ratio of small gaps to large gaps, with values close to 1 indicating a higher fre-



**Figure 3.** The observed gap size frequency distributions modeled as a power law function with  $\lambda = 2.6$ .

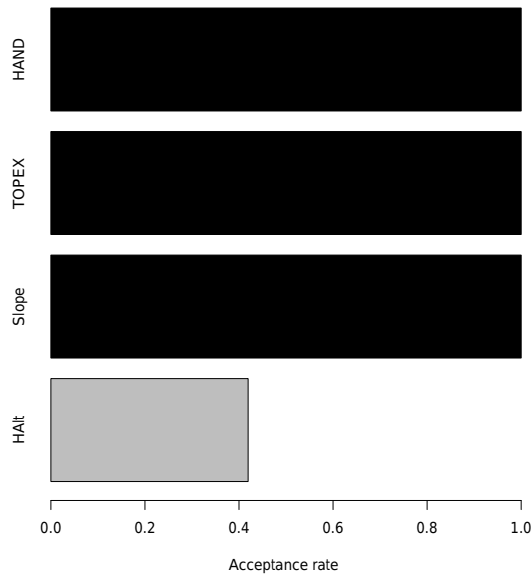
quency of large gaps and vice versa. Parameter estimates for slope and TRI show high occurrence of small gaps for large values of the two variables. Contrarily, the effects of DA, HAND, HALt, and TOPEX on  $\lambda$  are clearly negative, meaning that the frequency of large gaps increases with large values.

#### 3.4 The multivariate model

To define the final multivariate predictive model, we used the significant results of the univariate models together with the output of the PCA, in order to avoid multicollinearity.

##### 3.4.1 Variable selection

The first three PCA axes explained more than 80 % of the data variance. The first axis, which accounted for 36.45 % of the variance, was positively correlated with relative HALt and negatively correlated with HAND and DA, and thus clearly highlighted the local altitudinal gradient. The second axis explained an additional 28.5 % of variance and was positively correlated with the TRI and slope. The third axis explained a further 15.2 % of the variance and was correlated only with TOPEX (Fig. 2). The multivariate model was created using a Bayesian framework including four environmental variables: slope, TOPEX, HAND, and HALt, the explanatory variables that had an effect on  $\lambda$ . Finally, the KM methodological framework was used to select the most parsimonious model.



**Figure 4.** Results of the Kuo–Mallick algorithm for variable selection. Variables were included in the final model when their value was close to 100 %: slope, TOPEX and HAND.

Environmental covariates with posterior KM values close to 1, namely slope, TOPEX, and HAND (Eq. 9), were retained in the final model (Fig. 4). Parameter estimates of the final model indicated that the greatest effects on gap size distribution were caused by TOPEX and HAND.

## 4 Discussion

### 4.1 Methodology

#### 4.1.1 Gap detection

Delineating forest gaps is a persistent challenge for foresters and ecologists, among whom Brokaw’s gap definition (1982) has remained extensively used, in which “a ‘hole’ in the forest extending through all levels down to an average height of 2 m above the ground,” must be defined by an experienced observer. There are several studies that do not use this 2 m threshold definition of gaps, but instead 10 m (e.g., Hubbell et al., 1999; van der Meer and Bongers, 1996; Welden et al., 1991). However, in this study we have decided to use a probabilistic approach, modeling height distribution as a mixture of two normal laws. We found a height, 11 m, which is much higher than that in Brokaw’s definition but is consistent with our field experience, where woody debris, dead canopy tree boles, and residual saplings (i.e., remnants that survive the gap formation event) may rise well above 2 m. For example, (Hubbell et al., 1999) showed that small stems frequently remained in gaps up to 4–5 m in height, while (Lieberman et al., 1985) reported broken and damaged stems up to 10 m tall within a gap. The choice of the values of height and

threshold may be adapted to different forest types and topographic characteristics. In our case, the choice was fully data-driven using the DCM and the digital elevation model (DEM) and no ecological knowledge. Within our framework it is likely that in waterlogged areas, areas covered with mature trees that do not exceed the height thresholds may appear in our analysis as forest gaps. In order to clarify this question, an approach using time series would allow us to identify these “false” gaps that never get filled and thus are not part of the forest endogeneous dynamics. These are not gaps in the ecological meaning.

Defining minimum gap size is also a delicate proposition. Some authors, working with high-definition lidar data, have considered a minimum gap size ( $x_{\min}$ ) of 1 m<sup>2</sup> (Asner et al., 2013; Kellner and Asner, 2009). This minimum gap size is unrealistic from an ecological perspective given that a hole of several square meters in the canopy may simply reflect the distance between two crowns. Brokaw recommended a range from 20 to 40 m<sup>2</sup> based on his field experience. We have worked with a minimum gap size of 104 m<sup>2</sup>, and based this value on the minimized Kolmogorov–Smirnov distance between observed and predicted values.

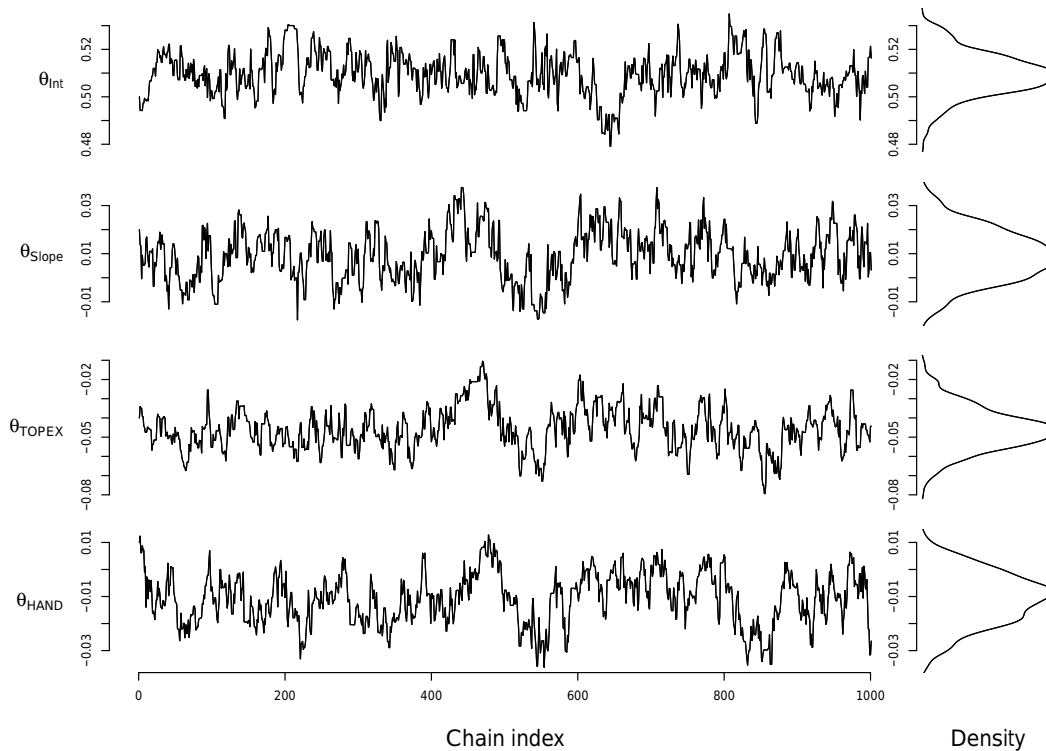
We built on previous studies that show that gap size distribution follows a power law distribution. However, the underlying mechanisms that control this distribution are still unclear. The Bayesian framework we developed allowed us to detail the contributions of each environmental variable to the size of each individual gap. Because the precise environmental variables were explicitly taken into account in the model likelihood of each gap, we were able to predict gap size distribution from environmental covariates, a difficult task when the scale exponent is estimated once, at the forest level, and compared between forests. The global scale exponent that we estimated for an average environment ( $\lambda = 2.6$ ) is consistent with some previous studies (Kellner and Asner, 2009; Kellner et al., 2011), though slightly larger than those of others: (Lobo and Dalling, 2014), [1.97; 2.15], and (Asner et al., 2013), [1.70; 2.03].

### 4.2 Environmental effects on gap size frequency distribution

For the first time, gap size distribution integrates environmental variables as a linear combination of the scale parameter ( $\lambda$ ) of a discrete Pareto distribution frequency. Our results suggest that three covariates drive the gap size frequency distribution in our forest: slope, HAND, and TOPEX (Fig. 5).

#### 4.2.1 Slope

Steep slopes are well known to directly impact tropical forest canopy structure (Bianchini et al., 2010). In this study, we found similar results to (Lobo and Dalling, 2014) on Barro Colorado Island (BCI); i.e., large gaps (smallest  $\lambda$ ) are more frequent on the gentle slopes. This may seem counter-



**Figure 5.** Posterior distribution of the environmental variables in the final multivariate model.

intuitive at first, as treefall may be (i) more prone to inducing cascading effects when slopes are steep and (ii) more frequent in slopes where soils are shallow with lateral drainage (Gourlet-Fleury et al., 2004), impeding deep rooting of trees. However, the forest turnover is more important in bottomlands where slopes are gentle (Durrieu de Madron, 1994). Considering that large gaps may be created solely by contiguous and independent treefalls, larger gaps may then be expected in bottomlands from a purely probabilistic approach. And given the positive link between wood density and steep slopes (Ferry et al., 2010), trees may be more resistant to cascading effects than they are in bottomlands.

#### 4.2.2 Water saturation

HAND is a binary variable that takes the value 1 on water-saturated soils. Because  $\lambda$  decreases when HAND equals 1, the frequency of large gaps increases in floodplains and bottomlands. These results support the findings of (Korning and Balslev, 1994), highlighting more dynamic forests in floodplains subject to large flooding events that lead to cascading treefall events. Together with Asner et al. (2013), our results suggest that we can effectively extend these results to bottomlands, where we already know that aboveground biomass and mean wood density are 10 % lower than on hilltops (Ferry et al., 2010). Given its ease of implementation on a land-surface model and its high predictive power, HAND

covariates present great potential applicability for gap size distribution prediction.

#### 4.2.3 Topographic exposure

The effect of topographic exposure on  $\lambda$  is consistent with our a priori hypothesis that wind-exposed areas would have a greater relative frequency of large gaps. Although hurricane damage does not occur in continental equatorial regions of the Amazon (Nelson et al., 1994), here we demonstrate that tree exposure has a large impact on gap size distribution. (Lobo and Dalling, 2014) observed no clear effect of TOPEX, and suggested that this index has a slight negative effect on gap size distribution. The results of this study are in line with the pioneering work of Negrón-Juárez et al. (2014), which showed that wind exposure is related to higher elevations that inflate the occurrence of larger gaps. However, coastal French Guianese forests exhibit different landscapes and landforms (Guitet et al., 2013). Our study area is made of dissected plateaus characterized by simple forms resembling hills (Guitet et al., 2013). It is possible that these characteristics, leading to unique combinations of landform elevations, may create complex terrain interactions that increase wind local speed and, in turn, cause large gaps. We conclude that topographic exposure is an appropriate index for predicting gap size distribution, but this must be confirmed in other landscape types.



## 5 Conclusions

To our knowledge, this is the first study where the precise environmental descriptors associated with each canopy gap were explicitly taken into account in the calculation of the model likelihood. We were able to do so because we wrote the general model likelihood as the product of all the single likelihoods (i.e., each gap had its own likelihood depending on the environmental covariate values). Doing so, we were able to predict gap size distribution from the fine environmental covariates, an impractical task when the scale exponent is estimated once at the forest level (i.e., mixing all the found gaps together) and compared between forests a posteriori. We also put forward an innovative method to define a height threshold and minimum gap size using two probabilistic approaches. The modeled distribution of canopy height as a mixture of two distributions provides a clear height threshold, while the minimization of KS distance between observed and predicted data proves to be efficient for setting the minimum gap size. We use a Bayesian framework in which the

model likelihood of each gap is expressed as a function of the unique environment local to the gap, highlighting the predominant role of the topographic exposure and waterlogging in determining gap size distribution. We expected that slope would also play an important role, with steeper slopes leading to larger gap sizes. However, we found that a steeper slope led to smaller gaps, as already highlighted by (Lobo and Dalling, 2014). We suggest that our modeling approach can be a basis for the development of large-scale methodologies using satellite data to understand gap-phase dynamics at a regional scale, combining lidar and radar remote sensing tools.

## 6 Data availability

Datasets used in this study are the property of the French National Forest Service, a private company. They are available upon request at [laurent.descroix@onf.fr](mailto:laurent.descroix@onf.fr).

## Appendix A

**Table A1.** List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the 0.0001th percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95 %)
Slope	Slope	°	0.119	[0.0416; 0.208]
Terrain ruggedness index	TRI	–	0.119	[0.083; 0.157]
TOPographic EXposure	TOPEX	–	–0.128	[–0.188; 0.00202]
Drained area	DA	m <sup>2</sup>	0.0843	[–0.0574; 0.179]
The hydraulic altitude	HAIt	m	–0.0135	[–0.04; 0.042]
HAND	HAND	–	–0.0615	[–0.152; 0.0162]

**Table A2.** List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the 0.001th percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95 %)
Slope	Slope	°	0.0735	[–0.02; 0.15]
Terrain ruggedness index	TRI	–	0.0718	[0.04; 0.10]
TOPographic EXposure	TOPEX	–	–0.082	[–0.12; –0.05]
Drained area	DA	m <sup>2</sup>	–0.0176	[–0.09; 0.05]
The hydraulic altitude	HAIt	m	–0.0177	[–0.05; 0.02]
HAND	HAND	–	–0.003	[–0.08; 0.09]

**Table A3.** List of environmental variables, abbreviations, units, and values of the posteriors in univariate models for a height threshold equal to the 0.01th percentile of the height distribution of the canopy.

Parameter	Abbreviation	Unit	Posterior value	Confidence interval (CI 95 %)
Slope	Slope	°	0.0975	[–0.02; 0.17]
Terrain ruggedness index	TRI	–	0.089	[0.05; 0.12]
TOPographic EXposure	TOPEX	–	–0.012	[–0.03; –0.32]
Drained area	DA	m <sup>2</sup>	–0.004	[–0.08; 0.05]
The hydraulic altitude	HAIt	m	0.063	[–0.04 ; 0.08]
HAND	HAND	–	–0.01	[–0.09; 0.06]

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We thank the handling editor, Ervan Rutishauser and Marijn Bauters for their thoughtful comments on a previous version of this paper. This study is part of the GFclim project funded by PO-Feder Région Guyane. Bruno Hérault was supported by a grant from the Investing for the Future program (managed by the 350 French National Research Agency – ANR, labex CEBA, ref. ANR-10-LABX-0025).

Edited by: A. Rammig

Reviewed by: E. Rutishauser and M. Bauters

## References

- Asner, G. P. and Mascaró, J.: Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric, *Remote Sens. Environ.*, 140, 614–624, doi:10.1016/j.rse.2013.09.023, 2014.
- Asner, G. P., Kellner, J. R., Kennedy-Bowdoin, T., Knapp, D. E., Anderson, C., and Martin, R. E.: Forest canopy gap distributions in the southern Peruvian Amazon, *PLoS one*, 8, e60875, doi:10.1371/journal.pone.0060875, 2013.
- Bianchini, E., Garcia, C. C., Pimenta, J. A., and Torezan, J.: Slope variation and population structure of tree species from different ecological groups in South Brazil, *An. Acad. Bras. Ciênc.*, 82, 643–652, 2010.
- Bicknell, J. E., Phelps, S. P., Davies, R. G., Mann, D. J., Struebig, M. J., and Davies, Z. G.: Dung beetles as indicators for rapid impact assessments: evaluating best practice forestry in the neotropics, *Ecol. Indic.*, 43, 154–161, 2014.
- Brokaw, N. V.: The definition of treefall gap and its effect on measures of forest dynamics, *Biotropica*, 14, 158–160, 1982.
- Brokaw, N. V.: Gap-phase regeneration in a tropical forest, *Ecology*, 66, 682–687, 1985.
- Chapman, L.: Assessing topographic exposure, *Meteorol. Appl.*, 7, 335–340, 2000.
- Clauset, A., Shalizi, C. R., and Newman, M. E.: Power-law distributions in empirical data, *SIAM Rev.*, 51, 661–703, 2009.
- Colson, F., Gond, V., Freycon, V., Bogaert, J., and Ceulemans, R.: Detecting natural canopy gaps in Amazonian rainforest, *Bois et forêts des tropiques*, 289, 69–79, 2006.
- Denslow, J. S., Ellison, A. M., and Sanford, R. E.: Treefall gap size effects on above- and below-ground processes in a tropical wet forest, *J. Ecol.*, 86, 597–609, 1998.
- Durrieu de Madron, L.: Mortalité des arbres en forêt primaire de Guyane française, *Bois et forêts des tropiques*, 239, 43–57, 1994.
- Espírito-Santo, F. D., Keller, M. M., Linder, E., Oliveira Junior, R. C., Pereira, C., and Oliveira, C. G.: Gap formation and carbon cycling in the Brazilian Amazon: measurement using high-resolution optical remote sensing and studies in large forest plots, *Plant Ecol. & Divers.*, 7, 305–318, 2014.
- Feeley, K. J., Davies, S. J., Ashton, P. S., Bunyavejchewin, S., Nur Supardi, M., Kassim, A. R., Tan, S., and Chave, J.: The role of gap phase processes in the biomass dynamics of tropical forests, *P. Roy. Soc. Lond. B Bio.*, 274, 2857–2864, doi:10.1098/rspb.2007.0954, 2007.
- Ferry, B., Morneau, F., Bontemps, J.-D., Blanc, L., and Freycon, V.: Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest, *J. Ecol.*, 98, 106–116, 2010.
- Fisher, J. I., Hurr, G. C., Thomas, R. Q., and Chambers, J. Q.: Clustered disturbances lead to bias in large-scale estimates based on forest sample plots, *Ecol. Lett.*, 11, 554–563, 2008.
- Frolking, S., Palace, M. W., Clark, D., Chambers, J. Q., Shugart, H., and Hurr, G. C.: Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure, *J. Geophys. Res.-Biogeo.*, 114, G00E02, doi:10.1029/2008JG000911, 2009.
- Gourlet-Fleury, S., Guehl, J.-M., and Laroussinie, O.: Ecology and management of a neotropical rainforest. Lessons drawn from Paracou, a long-term experimental research site in French Guiana, edited by: Gourlet-Fleury, S., Guehl, J. M., and Laroussinie, O., Elsevier, Paris, France, 326 pp., 2004.
- Guitet, S., Cornu, J.-F., Brunaux, O., Betbeder, J., Carozza, J.-M., and Richard-Hansen, C.: Landform and landscape mapping, French Guiana (South America), *Journal of Maps*, 9, 325–335, 2013.
- Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K., Condit, R., Wechsler, B., Wright, S. J., and De Lao, S. L.: Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest, *Science*, 283, 554–557, 1999.
- Kellner, J. R. and Asner, G. P.: Convergent structural responses of tropical forests to diverse disturbance regimes, *Ecol. Lett.*, 12, 887–897, 2009.
- Kellner, J. R., Asner, G. P., Vitousek, P. M., Twitten, M. A., Hotchkiss, S., and Chadwick, O. A.: Dependence of forest structure and dynamics on substrate age and ecosystem development, *Ecosystems*, 14, 1156–1167, 2011.
- Korning, J. and Balslev, H.: Growth and mortality of trees in Amazonian tropical rain forest in Ecuador, *J. Veg. Sci.*, 5, 77–86, 1994.
- Kuo, L. and Mallick, B.: Variable Selection for Regression Models, *The Indian Journal of Statistics, Series B (1960–2002)* 60, *Bayesian Analysis* (Apr. 1998), 65–81, published by: Indian Statistical Institute Stable, available at: <http://www.jstor.org/stable/25053023> (last access: 23 January 2017), 1998.
- Lieberman, D., Lieberman, M., Peralta, R., and Hartshorn, G. S.: Mortality patterns and stand turnover rates in a wet tropical forest in Costa Rica, *J. Ecol.*, 73, 915–924, 1985.
- Lloyd, J., Gloor, E. U., and Lewis, S. L.: Are the dynamics of tropical forests dominated by large and rare disturbance events?, *Ecol. Lett.*, 12, E19–E21, doi:10.1111/j.1461-0248.2009.01326.x, 2009.
- Lobo, E. and Dalling, J. W.: Spatial scale and sampling resolution affect measures of gap disturbance in a lowland tropical forest: implications for understanding forest regeneration and carbon storage, *P. Roy. Soc. Lond. B Bio.*, 281, 20133218, doi:10.1098/rspb.2013.3218, 2014.
- Mikita, T. and Klimánek, M.: Topographic Exposure and its Practical Applications, *Landscape Ecol.*, 3, 42–51, 2012.
- Molino, J.-F. and Sabatier, D.: Tree Diversity in Tropical Rain Forests: A Validation of the Intermediate Disturbance Hypothesis, *Science*, 294, 1702–1704, doi:10.1126/science.1060284, 2001.

- Negrón-Juárez, R. I., Chambers, J. Q., Hurtt, G. C., Annane, B., Cocke, S., Powell, M., Stott, M., Goosem, S., Metcalfe, D. J., and Saatchi, S. S.: Remote Sensing Assessment of Forest Disturbance across Complex Mountainous Terrain: The Pattern and Severity of Impacts of Tropical Cyclone Yasi on Australian Rainforests, *Remote Sensing*, 6, 5633–5649, 2014.
- Nelson, B. W., Kapos, V., Adams, J. B., Oliveira, W. J., and Braun, O. P.: Forest disturbance by large blowdowns in the Brazilian Amazon, *Ecology*, 75, 853–858, 1994.
- Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S.: Height above the nearest drainage—a hydrologically relevant new terrain model, *J. Hydrol.*, 404, 13–29, 2011.
- Puerta-Piñero, C., Muller-Landau, H. C., Calderón, O., and Wright, S. J.: Seed arrival in tropical forest tree fall gaps, *Ecology*, 94, 1552–1562, 2013.
- R Core Team: R: A language and environment for statistical computing, available at: <http://www.R-project.org> (last access: 21 January 2017), 2013.
- Riley, S. J.: Index That Quantifies Topographic Heterogeneity, *Intermountain Journal of Sciences*, 5, 23–27, 1999.
- Rüger, N., Huth, A., Hubbell, S. P., and Condit, R.: Response of recruitment to light availability across a tropical lowland rain forest community, *J. Ecol.*, 97, 1360–1368, 2009.
- Runkle, J. R.: Patterns of disturbance in some old-growth mesic forests of eastern North America, *Ecology*, 63, 1533–1546, 1982.
- SAGA, G.: System for automated geoscientific analyses, available at: [www.saga-gis.org/en/index.html](http://www.saga-gis.org/en/index.html) (last access: 21 January 2017), 2013.
- Schnitzer, S. A., van der Heijden, G., Mascaró, J., and Carson, W. P.: Lianas in gaps reduce carbon accumulation in a tropical forest, *Ecology*, 95, 3008–3017, 2014.
- van der Meer, P. J. and Bongers, F.: Formation and closure of canopy gaps in the rain forest at Nouragues, French Guiana 1996, 126, 167–179, doi:10.1007/BF00045602, blackboxCheck., 1996.
- Virkar, Y. and Clauset, A.: Power-law distributions in binned empirical data, *Ann. Appl. Stat.*, 8, 89–119, 2014.
- Wagner, F., Hérault, B., Stahl, C., Bonal, D., and Rossi, V.: Modeling water availability for trees in tropical forests, *Agr. Forest Meteorol.*, 151, 1202–1213, 2011.
- Welden, C. W., Hewett, S. W., Hubbell, S. P., and Foster, R. B.: Sapling Survival, Growth, and Recruitment: Relationship to Canopy Height in a Neotropical Forest, *Ecology*, 72, 35–50, doi:10.2307/1938900, 1991.
- Whitmore, T.: Canopy gaps and the two major groups of forest trees, *Ecology*, 70, 536–538, 1989.
- Yee, T. W.: The VGAM package for categorical data analysis, *J. Stat. Softw.*, 32, 1–34, 2010.