Spatial Statistics 13 (2015) 90–105

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Spatial pattern development of selective logging over several years

Salma Anwar *, Alfred Stein

Faculty of Geo-Information Science and Earth Observation, The University of Twente, Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 20 April 2014
Accepted 4 March 2015
Available online 21 May 2015

Keywords:
Selective logging
Point pattern
Inhomogeneous J-function
Geyer's saturation process
The area-interaction process

ABSTRACT

Selective logging gives currently a major contribution to ongoing deforestation in the Brazilian Amazonia. On satellite images, log-landing sites (LLS) are well visible, and they serve as a proxy to selective logging activities. In this study we analysed the spatial patterns of the LLS collected during the years 2000–2009 in a part of the Brazilian Amazonia, using spatial statistical methods. The purpose was to reveal important spatial and temporal characteristics of selective logging. After the spatial analysis, the patterns formed by the LLS were modelled using the higher-order Gibbs interaction models due to their suitability to model clustered patterns. The area-interaction model and Geyer's saturation model proved effective in modelling the clustered patterns in the absence of information about covariates. Results of both models conform closely to each other. We conclude that spatial statistical methods are powerful tools for analysing and interpreting the spatial patterns formed by selective logging.

1. Introduction

In the Brazilian Amazonian forests, selective logging for timber is a major source of forest degradation (Uhl et al., 1991). Adverse effects of selective logging on the Amazonian forests include damages to forest phenology (Koltunov et al., 2009), increasing vulnerability of a forest to fire (Gerwing, 2002; Fearnside, 2005), forest fragmentation (Broadbent et al., 2008) and wide spread of deforestation (Uhl

* Corresponding author.
E-mail address: anwar19631@utwente.nl (S. Anwar).

http://dx.doi.org/10.1016/j.spasta.2015.03.001
2211-6753/

et al., 1991; Fearnside, 2005). An important challenge faced by forest researchers is to detect and analyse the locations and patterns of forest perforations caused by selective logging for timber. Selective logging operations are almost impossible to monitor as much of the Amazonian region is inaccessible. Remote sensing, due to its synoptic view and fast coverage, may serve as a viable source for the purpose. Log-landing sites (LLS), the locations where the collected timber is stored, can be detected from the remote sensing images and may serve as a proxy for the selective logging activities in a surrounding area (Anwar and Stein, 2012).

There has been intensive work on analysing the deforestation patterns and processes in the Brazilian Amazonia. Efforts have been made to model the deforestation processes using different spatial analytical tools (Apan and Peterson, 1998; Alvis et al., 1998). Spatially explicit analysis and modelling of the locations of forest degradation caused by selective logging is, however, lacking. As logging operations vary in time, location and intensity (Matricardi et al., 2005), a spatial–temporal statistical analysis of the detected LLS may reveal important spatial and temporal characteristics of selective logging distribution. Due to small spatial extent of the LLS, it is natural to represent them as points in the maps derived from remote sensing images. Spatial point pattern statistics, then, may serve as appropriate tools for analysing and modelling the process that determines the LLS positions.

Different types of spatio-temporal data require different modelling approaches. The LLS maps generated from the remote sensing images are examples of the discrete-time spatio-temporal point process data. Discrete-time spatio-temporal point process data can arise in two ways; either the underlying process genuinely operates in discrete-time, or an underlying continuous-time process is observed at a discrete sequence of time-points (Diggle, 2005). LLS data may serve as an example of the latter as they are generated from the Landsat images acquired annually with almost the same time interval.

New logging activities grow from previously logged areas and thus extend more deeply into the interior core of remaining intact forest areas (Broadbent et al., 2008). LLS may thus exhibit interaction. Considering the interaction is important to understand the selective logging dynamics, whereas ignoring it can result in misleading conclusions about the spatial LLS distribution (Turner, 2009). If the LLS data are found to be of an inhomogeneous structure then different ecological and geographic factors may be responsible for the LLS distribution pattern. Modelling of such a nonstationary process is done using Monte Carlo Markov Chain (MCMC) methods. In the presence of interaction among LLS, such methods may become computationally extensive as they need to incorporate interaction terms as well as the spatially varying LLS intensity (Baddeley et al., 2000). When the LLS patterns exhibit interactions, pairwise interaction point processes are suitable choice. Pairwise interaction point processes are perhaps the most widely used sub-class of Markov point processes also known as Gibbs point processes. In a pairwise point process the configuration interacts only via pairs of points. Pairwise interaction models are good models for repulsive (regular) patterns but they do not sufficiently describe clustered patterns (Law et al., 2009). The lack of pairwise Markov models for clustered patterns led to the development of higher-order interaction processes such as Geyer's saturation and the area-interaction processes. These processes have infinite-order interactions and thus are well suited for modelling a clustered pattern.

The present study focuses on the assessment of spatial patterns formed by the LLS. Whereas the purely spatial aspects of selective logging have been dealt in Anwar and Stein (2015), the objective of this study is to discover important spatial–temporal characteristics of selective logging in the south-western part of the Brazilian Amazonia and then modelling the LLS patterns using a Gibbs model suitable for clustered patterns. By using patterns of several years, it is intended to better understand and quantify the spatial characteristics of the LLS distribution in the area and to reveal its temporal dynamics.

2. Study area and data description

The study area (Fig. 1) covering 30,000 km^2 is located in the south-western Brazilian Amazon. It constitutes the northern Rondônia state, north-western Mato Grosso state, and south-eastern Amazonas state — three of the five top timber-producing states which account for about 95% of the region's deforestation. The terrain of the region is undulated, ranging from 100 to 450 m above sea

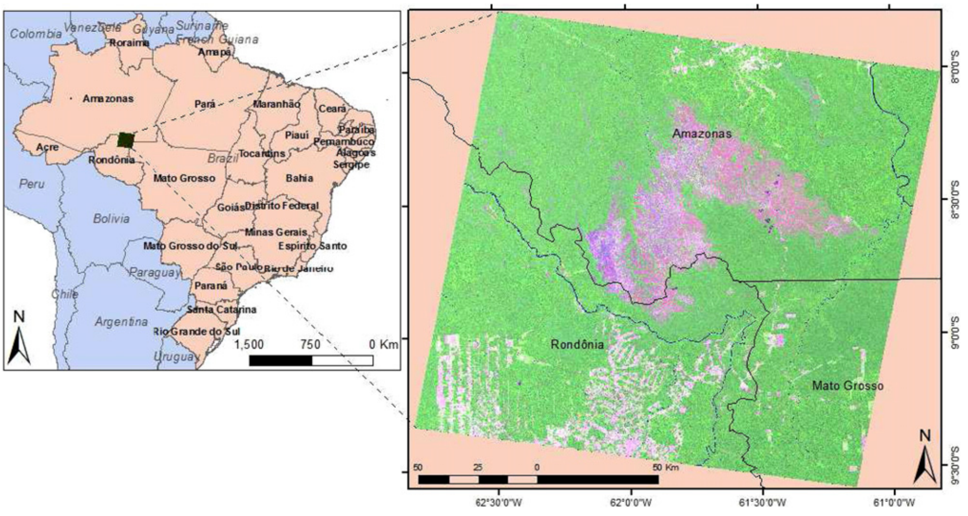

Fig. 1. Location of the study area in the north-west of Brazil and the Landsat-5 TM image corresponding the area.

level. Several soil types are found, mainly alifosols, oxisols, ultisols and alluvial soils. The region is drained by the Machado river, which is a tributary of the Madeira river (the biggest tributary of the Amazon). The climate in the area is classified as equatorial hot and humid, with tropical transition. It has a dry season that lasts from June to August. The annual average precipitation is 2016 mm and the annual average temperature is 25.5 °C (Lu et al., 2004).

The area is part of a colonisation project initiated by the Brazilian government in the 1970's. This project has played a major role in the deforestation of the area. The project is aimed at providing land to the landless people from the south-eastern region of Brazil. Migrants from the south converted forested areas into crop lands. The region has been deforested at high rates and has been attractive for timber loggers since then. The study area lies in the transition forest zone where commercially valuable timber trees are found in lower density (app. 20 m^3 ha^{-1}) than in the dense forests (35–40 m^3 ha^{-1}) (Monteiro et al., 2003). The loggers build roads in the area for carrying out mechanised logging operations (Lu et al., 2004). Along the roads, the logged timber is temporarily stored, usually at almost regular distances. From an LLS it is exported to the markets (Stone and Lefebvre, 1998; Souza and Barreto, 2000; Matricardi et al., 2005). These LLS may serve as spatial signature of areas under selective logging. Study of the LLS distribution patterns can be useful in determining the processes operating in the study area related to the logging activities. Moreover, from forestry context, the LLS can provide estimates of the locations and abundance of the valuable timber species in the study area.

Landsat 5-TM images of the study area for the years 2000, 2001, 2003, 2006, 2007, 2008 and 2009 with row 231 and path 66, were downloaded from the United States Geological Survey (USGS) website (http://earthexplorer.usgs.gov/). The images are processed to Standard Terrain Correction (Level 1T), which means that the images are radiometrically and geometrically corrected. The LLS were detected from the images by applying the methodology described in Anwar and Stein (2012) which is based on the linear spectral unmixing of the images. Three fraction images were calculated representing green vegetation, shade and soil, respectively. The soil fraction image was used to detect LLS within the forested areas.

3. Methodology

At the scale of the area covered by a Landsat image, the LLS can be considered as a spatial point pattern. The observed LLS data can then be described as a set of points (positions of LLS) $L = \{x_1, \ldots, x_n\}$ on a bounded region A in the plane, assumed to be a realisation of a spatial point process with as yet unknown properties.

3.1. Gibbs point process

Modelling LLS data tending to show interaction, requires the use of models that allow inclusion of the interactions. Gibbs processes are a natural choice in such case. The density of a Gibbs process (w.r.t. the unit rate Poisson process on A) is defined as

$$f(L) = a \prod_{z \subseteq L} \phi(z)$$

where a is a scaling factor, guaranteeing that f is a probability density, $\phi(z)$ denotes the interaction function. The interaction function describes the type and degree of interactions between LLS, and equals 1 in the absence of interaction.

3.1.1. Pairwise-interaction point processes

Pairwise interaction processes are the simplest form of Gibbs point processes. For the pairwise interaction processes, the conditional probability of finding a log-landing at location u given the complete realisation L of the process $\mathbf{L}$ is given by the conditional intensity function of the form:

$$\lambda(u, L) = \beta \prod_{i=1}^{n(L)} c(u, x_i)$$

where β is the first-order term and $c(u, v)$ is the second-order or pairwise interaction term of the interaction function ϕ, which measures dependence between LLS.

The pairwise interaction processes belong to the family of Markov (Gibbs) processes which are normally used to model the interaction structure of point patterns. These processes are adequate models for repulsive (regular) patterns but do not describe clustered patterns in a sufficient way (Grabarnik and Särkkä, 2001). An example is the Strauss model which was originally developed to model clustered patterns but later turned out to be suitable only for regular patterns. The conditional intensity function of the Strauss process is given by

$$\lambda(u, L) = \beta \gamma^{t(u,L)}$$

where $t(u, L) = s(L \cup u) - s(L)$, and $s(L)$ is the number of r- close pairs of distinct loglanding sites. β is related to the intensity of the loglanding process, whereas a value γ depicts the strength of interaction. If $\gamma = 1$ the model is a Poisson process with intensity β, whereas $0 < \gamma < 1$ means that a random LLS is less likely to occur at location u if there are many LLS in the neighbourhood. Thus it reflects inhibition in a LLS pattern. For $\gamma > 1$, the density becomes non-integrable when the number of points is not fixed, hence the density is defined only for $0 \leq \gamma \leq 1$, and fails to model the clustering nature of a LLS pattern.

To avoid the limitations of pairwise interaction point processes in modelling clustered patterns, higher-order interaction models such as Geyer's saturation model (Geyer, 1999) and the area-interaction model (Baddeley and Lieshout, 1995) have been developed. Note that a density of a higher-order interaction process depends on m- tuples of points, where $m > 2$.

3.2. Geyer's saturation point process

Geyer's saturation model is a modification of the Strauss process. It includes an additional parameter putting an upper bound s_b such that $t(x_i, L) \leq s_b$ is the contribution to the density of any single point x_i. It thus overcomes the normalising problem in the clustered case. Due to the upper bound, Geyer's saturation process is no more a pairwise interactions process, but a process of higher-order interactions. The density of the saturation process with interaction radius r and saturation threshold s_b can be written as:

$$f(L) = a\beta^{n(L)} \prod_{i=1}^{n(L)} \gamma^{\min\{s_b, t(x_i, L)\}}$$

with notations similar as in equation above for the Strauss process.

Fitting Geyer's saturation model involves estimation of two types of parameters: regular parameters and irregular parameters. Regular parameters like the intensity parameter β and the interaction γ can be estimated using the maximum likelihood or the pseudo-likelihood methods. Irregular parameters like the saturation threshold s_b and the interaction distance r, however, are difficult to estimate using the maximum likelihood method. These parameters can be estimated using prior-knowledge or by other methods e.g. by the profile pseudolikelihood which may or may not work well (Baddeley, 2010). For a nonstationary saturation process, the value β is replaced by a function $\beta(x)$ of a location x. If $s_b = 0$, the model reduces to the Poisson point process. If s_b is a finite positive number, then the model is valid for any positive value of interaction parameter γ (unlike the case of the Strauss process), with values $\gamma < 1$ describing an inhibitive pattern, and values $\gamma > 1$ describing a clustered pattern. If the saturation threshold s_b is very large, then $\min\{s_b, t(x_i, L)\}$ will be equal to $t(u_i, L)$. In such a case, Geyer's model reduces to the Strauss model (Turner, 2009). The interaction radius r of Geyer's saturation model is recommended to be a value in the interval from 0 to the maximum nearest neighbour distance (Baddeley and Turner, 2006).

3.3. The area-interaction point process

The area-interaction model is also a Markov point process with interactions of infinite order. The conditional intensity function of the area-interaction process is given by the equation

$$\lambda(u, L) = \beta\gamma^{-B(u,L)}$$

where $B(u, L) = A(L \cup u) - A(L)$ and $A(L \cup u)$ denotes the area formed by taking the union of the discs of radius r centred around all $x_i \in L$ and the location u, and $A(L)$ is the area formed by the discs of radius r centred at the $x_i \in L$. Thus $B(u, L)$ represents the area of that part of the disc of radius r centred at u that is not covered by discs of radius r centred at the other $x_i \in L$. As each LLS represents the area logged within a circle of radius r, we may interpret the conditional intensity $\lambda(u, L)$ as the conditional probability of having a new LLS at u, when other LLS's have known locations. $\gamma > 1$ gives a higher chance for the specification of such a new LLS with respect to a Poisson process with the intensity β; whereas for $\gamma < 1$ a new LLS is less likely to emerge within a small area around u when a set of locations of LLS is L, than in case of the Poisson process. Thus, $\gamma = 1$ corresponds to a Poisson process, $\gamma < 1$ produces a regular process, and $\gamma > 1$ a clustered process.

The area-interaction process can also be described in terms of the spatial birth-and-death processes (Baddeley and Lieshout, 1995). According to this definition, we may consider $\mathbf{L}$ as a LLS-generating process in which the existing LLS have exponential lifetimes, since the spatial signature of the LLS disappears within two to three years. Thus, a new LLS is determined at location u with rate $f(x \cup u)/f(x)$ related to the area accessible to u that is not already reached and logged from an existing LLS.

3.4. Model diagnostics

Various formal and informal methods to check the goodness-of-fit of a fitted point process model are available (Baddeley et al., 2008). Formal methods include χ^2 goodness-of-fit tests and Monte Carlo tests, whereas the informal methods include the residual analysis of a fitted point process model (Diggle, 2003). For Gibbs models, χ^2 goodness-of-fit test and Monte Carlo tests are theoretically not supported. Instead, the common distance-based summary functions are used to simulate the critical envelopes for the fitted models which are then used as tools for checking the adequacy of the fitted Gibbs models (Anwar et al., 2011; Myllymäki et al., 2013). Appropriateness of the inter-LLS interaction function in fitted Gibbs model is determined using an 'informal' validation tool known as Q–Q plot of the residuals (Baddeley, 2010). It compares empirical quantiles of the smoothed residual field to the corresponding expected empirical quantiles under the fitted model (Baddeley et al., 2005).

4. Results

We first analyse the spatial LLS data from 2006 and 2007, this is followed by the data from all the years when data were collected.

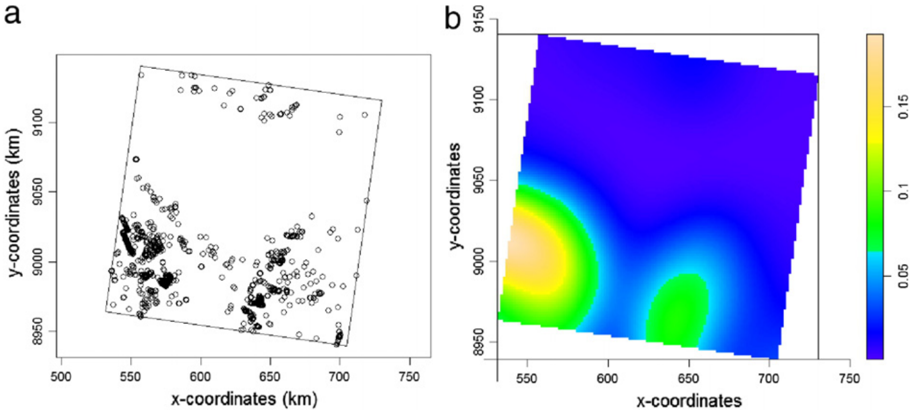

Fig. 2. (a) Loglanding sites detected in 2006, (b) kernel density estimate of intensity with kernel size equal to 20 km.

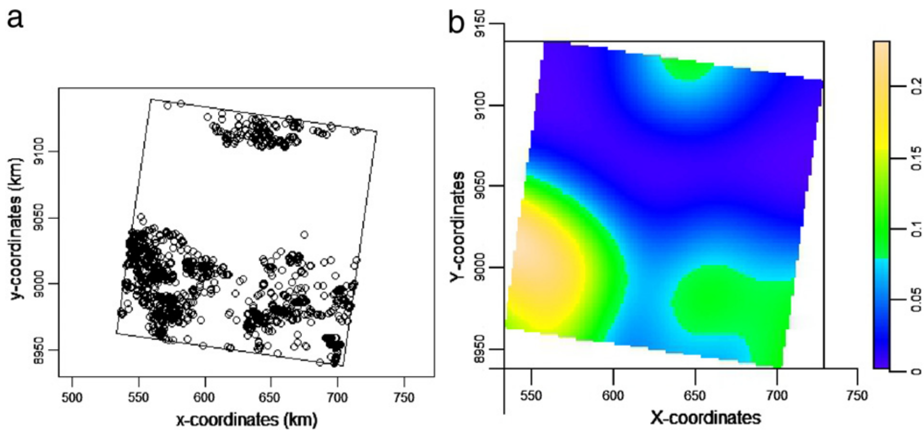

Fig. 3. (a) Loglanding sites detected in 2007, (b) kernel density estimate of intensity with kernel size equal to 20 km.

4.1. Intensity and J-function

Fig. 2(a) shows the locations of 990 LLS detected in year 2006. Gaussian kernel density estimate of the LLS intensity with kernel size equal to 20 km is presented in Fig. 2(b).

Fig. 3(a) shows the locations of 1829 LLS detected in 2007. Kernel density estimate of the LLS intensity with kernel size equal to 20 m is presented in Fig. 3(b).

To obtain a model that fits the LLS pattern well, we investigated the interaction structure of the LLS distribution. As Figs. 2(b) and 3(b) depict an inhomogeneous structure of the LLS distribution over the study area, we calculated the inhomogeneous J-function (Van Lieshout, 2011), a summary function suitable to assess the interaction structure which is intensity-reweighted moment stationary, for the LLS patterns of both years. Figs. 4 and 5 show the inhomogeneous J-functions for the LLS-2006 and LLS-2007 respectively along with the upper and lower envelopes corresponding to inhomogeneous Poisson point pattern. The observed inhomogeneous J-functions clearly lie below the simulated envelopes of the theoretical functions. This is evidence of a clustered pattern of the LLS in both years and we thus continue with Geyer's model and the area-interaction model.

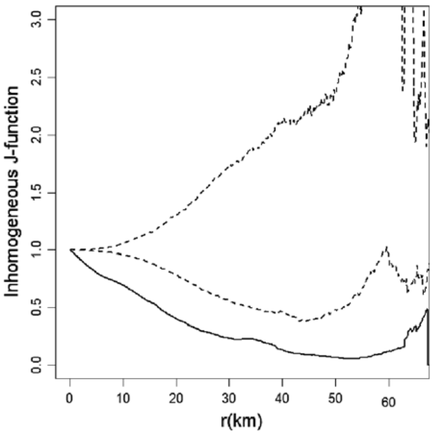

Fig. 4. Inhomogeneous J-function calculated using kernel density estimate with kernel size $=$ 20 for LLS-2006 data. The solid curve shows the inhomogeneous J-function for the LLS-2006 data, whereas the dotted curves are the upper and lower envelopes of the same function calculated from simulations of corresponding inhomogeneous Poisson process.

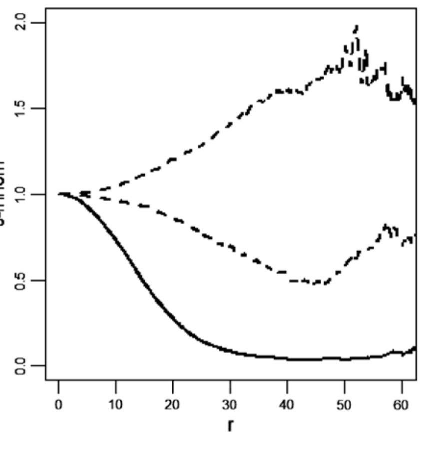

Fig. 5. Inhomogeneous J-function calculated using kernel density estimate with kernel size $=$ 20 for LLS-2007 data. The solid curve shows the inhomogeneous J-function for the LLS-2007 data, whereas the dotted curves are the upper and lower envelopes of the same function calculated from simulations of corresponding inhomogeneous Poisson process.

4.2. Geyer's saturation model for the LLS data

4.2.1. Modelling LLS data of years 2006 and 2007

To model the clustered pattern of LLS-2006, we first fitted Geyer's saturation model due to its ability to take into account the clustering nature of a pattern. For fitting this model we need to specify the irregular parameters, based upon the pseudo-likelihood estimation technique. The pseudo-likelihood estimates of these parameters were calculated and the profile log pseudo-likelihood was plotted in Fig. 6.

From Fig. 6 we observe that the log-pseudolikelihood attains its maximum value for $r = 5$ and $s = 14$. To test how well the model with these estimated values of parameters fits the LLS data, we calculated and plotted the inhomogeneous K-summary function (Baddeley and Turner, 2005; Baddeley et al., 2000) for the LLS data along with the envelopes of the same summary function calculated for 99 simulated realisations of the fitted model in Fig. 7.

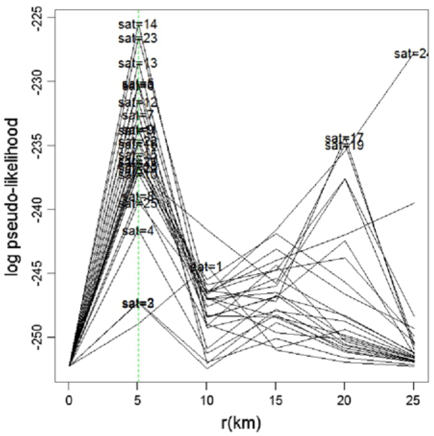

Fig. 6. Profile-pseudolikelihood for LLS-2006.

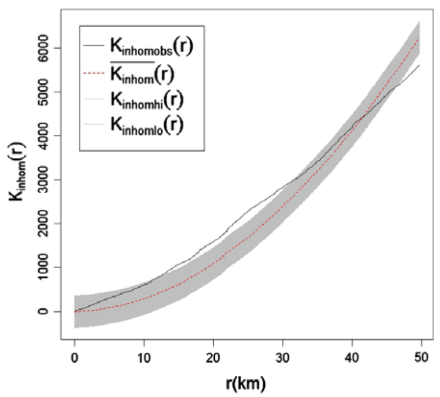

Fig. 7. Envelopes of the inhomogeneous K-function calculated for 99 simulated realisations of fitted Geyer's saturation model with parameters $s_b = 15$ and $r = 5$, along with the curve of the inhomogeneous K-function calculated for the LLS-2006 data.

Fig. 7 shows a substantial lack of fit particularly within distance between 10 and 30 km of a typical LLS of our data set with the K-function clearly outside the range of values of the K-function for simulated patterns. The plot suggests that the LLS are more clustered within this range than suggested by our choice of parameters. Hence $r = 5$ and $s_b = 14$ do not serve as optimal choices for Geyer's model to the 2006 data and we conclude that log-pseudolikelihood estimation does not work well for those data. In such a case, the alternative strategy to estimate the parameter values is to consider all possible values of 'r' and 's_b' within a reasonable range and their combinations for fitting the model and then testing goodness-of-fit of all the models fitted thus to select the best fitting model for the LLS data. For this purpose we defined the range of possible values for the two parameters. Fig. 4 shows that the effective range of r does not go beyond $r = 30$ km. As its value increases beyond its effective range, the simulated envelopes span over a wider range and the relative noise in the simulated envelopes also increases. Hence the range of r can be inferred from the inhomogeneous J-functions. Moreover, we calculated the maximum nearest neighbour distance in the LLS pattern which was found to be equal to 28 km. Therefore, we considered values of r from 5 to 25 km at steps of 5 km for fitting Geyer's saturation model. For the saturation threshold s_b, low values are preferred since these make the evaluation of the log-pseudolikelihood computationally fast. We considered all values of s_b from 5 to 20 at steps of 5 and Geyer's model was fitted to the LLS data using all combinations of r and s_b within these ranges. To decide upon the form of spatial trend which is most suitable to

Table 1
Estimated values of the interaction parameter γ with respective AIC values of Geyer's saturation model calculated for different values of interaction radius r and saturation threshold s_b for the LLS-2006.

Interaction radius r	Saturation threshold s_b	AIC	Interaction parameter γ
5	5	5400	1.40
10	5	4174	1.23
15	5	2090	1.49
20	5	1064	>10
25	5	517	>10
5	10	5185	1.21
10	10	4110	1.12
15	10	2100	1.16
20	10	1062	1.67
25	10	516	>10
5	15	4954	1.16
10	15	4028	1.10
15	15	2117	1.06
20	15	1044	1.70
25	15	516	>10
5	20	4819	1.13
10	20	3897	1.10
15	20	2109	1.04
20	20	1076	1.10
25	20	514	>10

capture the remaining variability in the LLS distribution, AIC values were calculated using the log-pseudolikelihood for different forms of trend functions and compared. The intensity function which is log-linear in the cartesian coordinates was found to be the best choice.

Table 1 shows the estimated parameters of fitted intensities for the different models along with their AIC values, using the intensity function that is log-linear in the cartesian coordinates.

From Table 1 we first observe that the value of the interaction parameter γ decreases and tends to one if the value of s_b increases and r is kept constant, i.e. the model reduces to a Poisson process. This is because γ indicates the probability of the number of LLS neighbours greater than or equal to the saturation threshold s_b. In other words, keeping the values of r constant, the value of γ decreases for large values of s_b because there is a lower chance then for a LLS to have more than 's_b' neighbours. In that case we cannot determine the true intensity of clustering in a particular data set. Therefore the value of s_b is kept to a minimum so as to maintain the effectiveness of Geyer's model in accounting for a clustered nature of the LLS data and describing the strength of its clustering in numerical terms. Secondly, we observe that as the value of r increases to the maximum nearest-neighbour distance in the LLS pattern, which in this case was equal to 28 km, γ takes large values as compared to the previous ones, indicating the presence of strong clustering. Such values of γ do not appear to be credible and could be anomalies or numerical artefacts caused by inaccuracies of pseudo-likelihood method, and therefore were disregarded.

Next, to find the optimal values of r and s_b, we performed a goodness-of-fit test of the above fitted models. For that purpose, the inhomogeneous K-summary functions were calculated along with the envelopes of the same function for 99 realisations of the fitted models. The inhomogeneous $K(u)$-function determines the expected total weight of all random LLS within a distance u of a typical LLS of the process (Baddeley et al., 2000). From the simulated envelopes, we found that the model with log-linear trend and the values $r = 15$ km and $s_b = 5$ are adequate choices to represent the spatial distribution of the LLS. For this model, γ was estimated to be equal to 1.49 thus providing an evidence of moderate clustering in the LLS distribution. The K-function for the model and its critical envelopes are plotted in Fig. 8.

From Fig. 8 we observe that the inhomogeneous K-function of the data set calculated over a range of interaction distance lies well within the critical envelopes of the summary function under the fitted saturation model. We thus conclude that the our chosen model is a suitable fit to the LLS-2006.

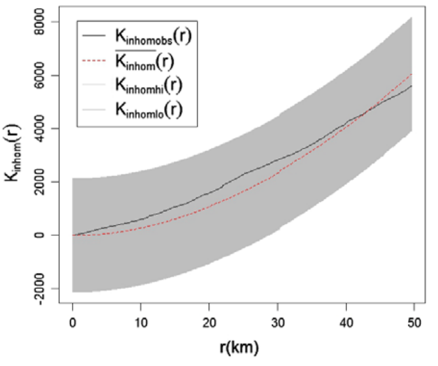

Fig. 8. Envelopes of the inhomogeneous K-function calculated for 99 simulated realisations of fitted Geyer's saturation model with parameters $s_b = 5$ and $r = 15$, along with the curve of the inhomogeneous K-function calculated for the LLS-2006 data.

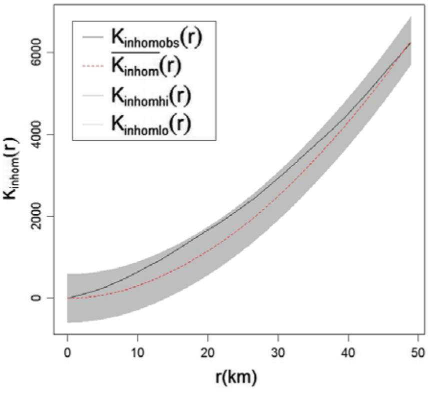

Fig. 9. Envelopes of the inhomogeneous K-function calculated for 99 simulated realisations of fitted Geyer's saturation model with parameters $s_b = 5$ and $r = 15$, along with the curve of the inhomogeneous K-function calculated for the LLS-2007 data.

Following the methodology adopted for modelling the LLS data for year 2006, we inferred the effective range of $r = 20$ km for Geyer's saturation model from Fig. 5. Table 2 shows estimated parameters of the log-linear intensity functions for Geyer's model with different values of interaction parameter and saturation threshold.

To find the optimal values of r and s_b, we performed goodness-of-fit test of all the fitted models above. For this, we calculated the inhomogeneous K-summary functions along with their envelopes for 39 simulated realisations of the fitted models. From the simulated envelopes, we found that the model with a log-linear trend and the values $r = 15$ km and $s_b = 5$ is a good choice to represent the LLS distribution. For this model, interaction parameter is equal to 1.90 which provides an evidence of clustering in the LLS distribution. The corresponding K-function and its critical envelopes are plotted in Fig. 8.

From Fig. 9 we observe that the inhomogeneous K-function of LLS-2007 data lies well within the critical envelopes of the summary function under the fitted saturation model, therefore we conclude that our chosen model is a suitable fit to the LLS-2007 data.

4.3. Modelling the LLS data of years 2000–2009

The above described analysis and modelling approach was adopted for the LLS data of all the years between 2000 and 2009 for which we had data. Table 3 provides a summary of the final fitted models.

Table 2
Estimated values of the interaction parameter γ with respective AIC values of Geyer's saturation model for different values of interaction radius r and saturation threshold s_b for the LLS-2007.

Interaction radius r	Saturation threshold s_b	AIC	Interaction parameter γ
5	5	5400	1.48
10	5	4174	1.63
15	5	2090	1.90
20	5	1064	>10
5	10	5185	1.24
10	10	4110	1.30
15	10	2100	1.08
20	10	1062	>10
5	15	4954	1.15
10	15	4028	1.18
15	15	2117	1.15
20	15	1044	>10
5	20	4819	1.13
10	20	3897	1.13
15	20	2109	1.14
20	20	1076	>10

Table 3
Estimated values of the interaction parameter γ with respective AIC values of Geyer's saturation model calculated for different values of interaction radius r and saturation threshold s_b for the LLS data of the years between 2000 and 2009 for which data were available.

Year	Number of LLS	Interaction radius r	Saturation threshold s	AIC	Interaction parameter γ
2000	650	5	5	3531	1.52
2001	917	10	5	2687	1.59
2003	752	5	5	4140	1.45
2006	990	15	5	2090	1.50
2007	1829	15	5	3509	1.90
2008	1312	20	5	1940	2.09
2009	2301	15	5	4215	1.65

Table 3 shows that the LSS intensity, as manifested by the number of LLS, is lower between 2000 and 2006 as compared to the successive years. It reflects that selective logging activities increased after 2006. In 2007 the number of LLS almost doubled as compared to the number of LLS in 2006. The number of LLS remains high in the latter years as well. Effect of the increasing intensity can be seen in the estimates of the interaction radius for different years and the associated interaction parameters. For 2000–2003, the estimated interaction radii and the interaction parameters are smaller as compared to the ones for years 2006–2009. This means that for the years 2000–2003, log-landing activities were mainly confined to smaller areas with less inter-LLS interactions i.e. the smaller interaction radii provide evidence of smaller cluster sizes of LLS for the given years. Smaller interaction parameters with smaller interaction radii prove that the LLS are less tightly clustered over regions of smaller sizes as compared to the later years when we find that the interaction radii increase as well as the interaction parameters, indicating that the LLS are largely and more tightly clustered. For each year, the LLS detected in the previous years and repeatedly detected in the given year were taken out to restrict the analysis for a given year only to the new LLS detected in that year. Since for the years 2003 and 2006 we do not have data available for all the previous years, results for those years must be interpreted with care keeping in mind that the LLS data for these years may include some data from the previous years as well.

4.4. The area-interaction model

For the area-interaction model, the pseudo-likelihood profile could not be calculated using Spatstat software (Baddeley and Turner, 2005) because of the memory allocation problem. Therefore, to

Table 4
Estimated values of the interaction parameter γ with respective AIC values of the area-interaction model calculated for different values of interaction radius r for the loglanding sites data of year 2006.

Interaction radius r	AIC	Interaction parameter γ
1	5040	3.87
2	4602	1.50
3	4652	1.22
4	4726	1.14
5	4771	1.09
6	4597	1.07
7	4411	1.06
8	4197	1.05
9	4034	1.04
10	3856	1.04

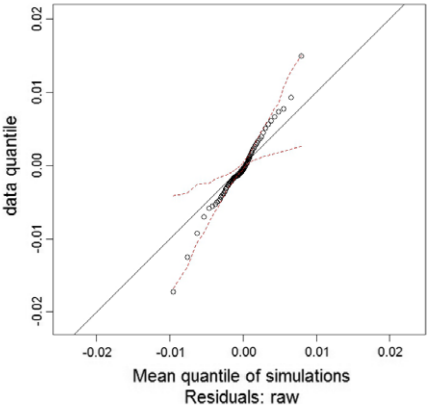

Fig. 10. Q–Q-plot for the area-interaction model for LLS-2006, using $r = 2$. The dotted lines show the envelopes of the inhomogeneous K-function for 19 simulated realisations of the fitted model.

estimate the optimal value of r of the area-interaction process, we fitted the area-interaction model to the LLS data using r values of 1 to 10 km with a step of 1 km and tested goodness-of-fit. For each value of r we used a different forms of the intensity function and compared AIC values of the models. The intensity function that is log-linear in the cartesian coordinates was the best choice since the higher-order polynomials did not result in significant improvement in the AIC value, and quadratic or cubic terms in the polynomials were all very close to zero. Moreover values of the interaction parameter η remained almost unaffected when including higher-order polynomial terms. Table 4 shows the AIC values and the estimated interaction parameter γ for all different values of r using the intensity function that is log-linear in the Cartesian coordinates.

From the table we observe that the value of γ decreases and tends to one as r increases. This means that the LLS-generating process tends to be a Poisson process, lacking interaction between the LLS, as the influence zone of each LLS increases. To determine which combination of r and γ best represents the LLS distribution, we checked the goodness-of-fit of the models for all different values of γ using the inhomogeneous K-function, including the envelopes of 19 realisations of the fitted models. From the plots, we observed that all the K-functions calculated from the observed patterns lied well within the critical envelopes. In such a situation it was difficult to determine the optimal value of r or the best fitting model. Therefore, we investigated the residuals from the fitted models using the Q–Q-plot, being a commonly used plot for Gibbs models (Turner, 2009; Baddeley et al., 2005). We found that the Q–Q-plot (Fig. 10) for $r = 2$ km was showing the best fit, although still not the perfect one, for the LLS-2006.

Table 5
Estimated values of the interaction parameter γ for different values of interaction radius r of the area-interaction model fitted to LLS-2007.

Interaction radius r	AIC	Interaction parameter γ
1	7947	4.02
2	7124	1.53
3	7335	1.25
4	7243	1.17
5	7123	1.14
6	4597	1.12
7	4411	1.10

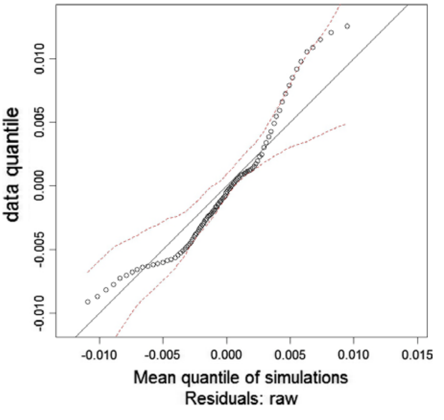

Fig. 11. Q–Q-plot for the area-interaction model for LLS-2007 data, using $r = 2$. The dotted lines show the envelopes of the inhomogeneous K-function for 19 simulated realisations of the fitted model.

Next, we fitted the area-interaction model to the LLS-2007 data for values of r from 1 to 10 km with a step of 1 km. For each value of r we used different forms of spatial trend. We compared AIC values of the models to determine the form of spatial trend most suitable to represent the variability in LLS data. The intensity function which is log-linear function of the cartesian coordinates was found to be the best choice. Table 5 shows the AIC values and the estimated interaction parameter γ for all different values of r using the log-linear intensity function.

To select the model that best represents the LLS spatial distribution, we checked goodness-of-fit of the models. The inhomogeneous K-functions were calculated along with the envelopes of the same function for 19 realisations of the fitted models. For this year also, the plots of the functions showed that all the calculated K-functions for the data lied well within the critical envelopes simulated from the fitted models. Therefore the simulation envelopes of the fitted models failed to provide any clues about the best fitting model. To further assess suitability of the models we investigated the residuals from all the fitted models. For this, Q–Q-plots of the residuals were used to determine the most optimal value of r and the corresponding best fitting model. We found that the Q–Q-plot (Fig. 11) for $r = 2$ provided the best fit, although still with hint of some misspecification, for the LLS-2007.

The analysis and modelling technique described above was adopted for LLS data of all the years available. Table 6 provides a summary of the final fitted area-interaction models for the years between 2000 and 2009.

From the table we observe that interaction parameters for all the years are significantly larger than 1, providing evidence of clustering in the LLS distribution of all the years. However we may note that for the years 2000–2003, the interaction radii are larger than the interaction radii for the years 2006–2009. Moreover, the table points to the two important features of the data set: the interaction distances are smaller and the interaction parameters of the area-interaction model are larger for the years after 2006 as compared to the previous years. This means that in the years 2006 and prior, the

Table 6
Estimated values of the interaction parameter γ with respective AIC values of the area-interaction model calculated for different values of interaction radius r for the LLS data of year 2000–2009.

Year	Number of LLS	Interaction radius r	Interaction parameter γ
2000	650	3	1.25
2001	917	3	1.17
2003	752	3	1.25
2006	990	2	1.50
2007	1829	2	1.53
2008	1312	2	1.46
2009	2301	2	1.51

LLS are located at the larger distances from each other and since the interaction parameters are smaller we cannot expect many LLS within a given interaction radius around a point in the pattern. Whereas in the later years, smaller interaction distances indicate that the points are located closer to each other and even within smaller distances we can expect larger number of LLS around a chosen LLS. This leads to the conclusion that the LLS data are more sharply and tightly clustered in the years after 2006. Hence the mechanism of LLS generation has changed, which may indicate a change of deforestation policies.

5. Discussion

Spatial point pattern statistics have been used to characterise spatial point patterns in forestry science. To the best of our knowledge, such models have not been applied to describe the patterns formed by spatial signatures left after selective logging of forests for timber. In this paper and in Anwar and Stein (2015), we have explored point pattern techniques to characterise the LLS patterns.

We applied the area-interaction model and Geyer's saturation process to investigate the spatial distribution of LLS. The Strauss process and the area-interaction process are examples of point processes from two large classes of processes: the distance-interaction processes and the shot-noise weighted processes (Van Lieshout and Molchanov, 1998). We have shown the effectiveness of both types of processes in modelling clustered patterns such as the LLS distribution in our study area. In general, both approaches worked well for our data and the results of both models are close to each other. Simulation envelopes and a Q–Q-plot were used as diagnostics for model validation. Although the Q–Q-plots showed that model fitting was satisfactory, they did not show a perfect fit. This does not necessarily mean that the interaction structure was mis-specified, as the discrepancies in a Q–Q-plot may result from a number of uncontrolled factors such as mis-specification in the trend function or presence of outliers. It was also observed in this study that the Q–Q-plot is sensitive to slight mis-specifications in the range of interpoint interactions. Therefore different values of interaction distance with small intervals must be tested. This, however, may not always be feasible due to large number of possible values within a given range and the large amount of time it requires to calculate Q–Q-plot for each value.

Correct interpretation of Q–Q-plot also depends upon the correct form of spatial trend and intensity function used. In case of an inhomogeneous LLS pattern, even if the spatial trend is correctly determined, the intensity function may involve dependence on covariates. Such dependence may or may not be observable and it may thus be difficult to measure and incorporate all of the covariate effects into a fitted model if dependence on some or all of the covariates is not observable. Hence the Q–Q-plot may show lack of fit resulting in misleading conclusions about the fitted model. Specifically, the distribution of LSS is believed to be determined by a number of local and regional socio-economic and environmental factors. At the scale of a single Landsat scene, it is not possible to measure the effect of all the factors involved and hence the intensity may not be fully modelled in terms of covariates.

In practice, Gibbs models currently available for use in data analysis are low in number and limited in scope. In real world situations, however, we come across a plethora of different examples of point patterns with different distributional characteristics. The available set of spatial statistical models is therefore insufficient to represent this diversity. It cannot be expected that one or other of the developed models will appropriately apply to a given situation. Due to diversity of real-world situations, the scope of available set of models might need to be expanded. Limitations in doing so concern the formulation itself of such a model and fitting it to the data set may involve skillful programming, apart from the more technical and theoretical prerequisites such as integrability and local stability of the functions used. Moreover distinguishing between existing models is sometimes difficult, in particular as only the data and not detailed information of the process involved in the data generation is available to guide the choice of suitable model.

Despite the limitations stated above, our fitted models proved to be satisfactory to model the distributional characteristics of the LLS data, in particular from the diagnostic tools used. In a forestry context, our study can provide us with an overview of the characteristics of LLS data in spatially explicit and quantitative terms. Our study has presented several spatial point pattern methods which are, we believe, of importance to the forest statistician. The LLS distribution reflects the location and density of valuable timber species in a study area and may point to possible degradation. Forest degradation caused by selective logging, however, is a complex process affected by a plethora of socio-economic and environmental factors. With the use of covariate information, it is expected that the fitted spatial statistical models can give us improved results which can be used for predictive purposes to determine the future trends in the distribution of selective logging. Further, the change in parameter values over the years has given some idea in changes in deforestation activities. The stronger clustering points to an intensified deforestation, where new LLS are located at shorter distances than before. Such information is useful for forestry studies and may be lucrative for timber loggers in order to carry out selective logging operations in a cost- and time-effective manner.

In the future, the results of this study could be used to investigate the forest structure left after the logging operations. This information may also be required by the forest conservation agencies to plan effective strategies for exploiting the forest resources without causing significant damage to its canopy structure and composition of its flora and fauna. The spatial point pattern methods may be included into the everyday toolbox of exploratory methods for analysing the selective logging distributions in the Amazonian forests.

6. Conclusions

The spatial statistical methods presented in the paper provide a useful tool to get insight into the important spatial and temporal characteristics of LLS that are used as a proxy for deforestation activities in the Brazilian Amazonia. Plots of LLS reflected inhomogeneity, and the inhomogeneous J-function were used to analyse their spatial distribution. The J-function helped to infer the ranges and types of interaction using the non-parametric form of the intensity function and showed that selective logging operations are strongly aggregated. Geyer's saturation process and the area-interaction process were shown to be effective in modelling a clustered pattern such as selective logging distribution in the study area.

This study leads to the following conclusions that are relevant for studies on spatial statistical analysis of forest degradation caused by selective logging in Amazonian forest.

1. A spatial statistical analysis was useful to understand and interpret the LLS pattern as it occurred on a Landsat image. It showed non-stationarity and clustering in their spatial pattern that helped to determine a suitable model to represent their distribution. 2. Geyer's saturation model and the area-interaction process model serve as suitable choices to model the clustered pattern of the LLS distribution even in the absence of any information about covariates information. 3. Results of Geyer's saturation process model and the area-interaction process model conform closely to each other hence both models prove to be effective in representing the clustered LLS pattern.

References

Alvis, D.S., Pereira, J.L.G., De Souza, C.L., Yamaguchi, F., 1998. Characterizing landscape changes in central Rondonia using Landsat TM imagery. Int. J. Remote. Sens. 20, 2877–2882.

Anwar, S., Stein, A., van Genderen, J., 2011. Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks. In: Shi, W. (Ed.), Advances in Geo-spatial Information Science. In: ISPRS Book Series, Taylor & Francis, London, pp. 125–140.

Anwar, S., Stein, A., 2012. Detection and spatial analysis of selective logging with geometrically corrected Landsat images. Int. J. Remote. Sens. 33, 7820–7843.
Anwar, S., Stein, A., 2015. Use of spatial statistics to investigate early forest degradation activities as detected from satellite images. Spat. Stat. 12, 50–64.
Apan, A.A., Peterson, J.A., 1998. Probing tropical deforestation: The use of GIS and statistical analysis of georeferenced data. Appl. Geogr. 18, 137–152.
Baddeley, A.J., Lieshout, M.N.M., 1995. Area-interaction point processes. Ann. Inst. Statist. Math. 47, 601–619.
Baddeley, A., Turner, R., Møller, J., Hazelton, M., 2005. Residual analysis for spatial point processes. J. R. Stat. Soc. Ser. B 67, 617–666.
Baddeley, A., Turner, R., 2005. Spatstat: an R package for analyzing spatial point patterns. J. Stat. Soft. 12 (6), 1–42. URL: www.jstatsoft.org.
Baddeley, A.J., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statist. Neerlandica 54, 329–350.
Baddeley, A.J., Turner, R., 2006. Modelling spatial point patterns in R. In: Case studies in spatial point pattern modelling. In: Lecture notes in Statistics, vol. 185. Springer, London, pp. 23–74.
Baddeley, A.J., Møller, J., Pakes, A.G., 2008. Properties of residuals for spatial point processes. Ann. Inst. Stat. Math. 60, 627–649.
Baddeley, A.J., 2010. Modelling strategies. In: Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P. (Eds.), Handbook of Spatial Statistics, Chapter 12. CRC Press, Boca Raton, pp. 371–402.
Broadbent, E.N., Asner, G.P., Keller, M., Knapp, D.E., Oliveira, P.J.C., Silva, J.N., 2008. Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. Biol. Conserv. 141, 1745–1757.
Diggle, P.J., 2003. Statistical analysis of spatial point patterns. Arnold Publishers, London.
Diggle, P.J., (2005) Spatio-temporal point processes: Methods and applications. Johns Hopkins University, Dept. of Biostatistics Working Papers. Working Paper 78. http://biostats.bepress.com/jhubiostat/paper78.
Fearnside, P.M., 2005. Deforestation in Brazilian Amazonia: History, rates, and consequences. Conserv. Biol. 19, 680–688.
Gerwing, J.J., 2002. Degradation of forests through logging and fire in the eastern Brazilian Amazon. Forest. Ecol. Manag. 157, 131–141.
Geyer, C.J., 1999. Likelihood inference for spatial point proceses. In: Barndorff-Nielsen, O.E., Kendall, W.S., Van Lieshout, M.N.M. (Eds.), Stochastic Geometry: Likelihood and Computation. In: Monographs on Statistics and Applied Probability, vol. 80. Chapman and Hall, London, pp. 79–140.
Grabarnik, S., Särkkä, A., 2001. Interacting neighbour point processes: some models for clustering. J. Statiat. Comput. Simul. 68 (2), 103–125.
Koltunov, A., Ustin, S.L., Asner, G.P., Fung, I., 2009. Selective logging changes forest phenology in the Brazilian Amazon: Evidence from MODIS images time series analysis. Remote. Sens. Environ. 113, 2431–2440.
Law, R., Illian, J., Burslem, D.F.R.P., Gratzer, G., Gunatilleke, C.V.S., Gunatilleke IAUN, 2009. Ecological information from spatial patterns of plants: Insights from point process theory. J. Ecol. 97, 616–628.
Lu, D., Batistella, M., Moran, E., Mausel, P., 2004. Application of spectral mixture analysis to Amazonian land-use and land-cover classification. Int. J. Remote. Sens. 25, 5345–5358.
Matricardi, E.A.T., Skole, D.L., Cochrane, M.A., Qi, J., Chomentowski, W., 2005. Monitoring selective logging in tropical evergreen forests using Landsat: multitemporal regional analyses in Mato Grosso. Brazil. Earth Interac. 9, 1–24.
Monteiro, A.L., Souza Jr, C.M., Barreto, P., 2003. Detection of logging in Amazonian transition forests using spectral mixture models. Int. J. Remote. Sens. 24, 151–159.
Myllymäki, M., Mrkvicka, T., Seijo, H., Grabarnik, P., 2013. Global envelope tests for spatial processes. preprint arXiv:1307.0239.
Souza, C., Barreto, P., 2000. An alternative approach for detecting and monitoring selectively logged forests in the Amazon. Int. J. Remote Sens. 21, 173–179.
Stone, T.A., Lefebvre, P., 1998. Using multi-temporal satellite data to evaluate selective logging in Para, Brazil. Int. J. Remote Sens. 19, 2517–2526.
Turner, R., 2009. Point patterns of forest fire locations. Environ. Ecol. Stat. 16, 297–223.
Uhl, C., Verissimo, A., Mattos, M.M., Brandino, Z., Vieira, I.C.G., 1991. Social, economic, and ecological consequences of selective logging in an Amazon frontier: the case of Tailândia. Forest. Ecol. Manag. 46, 243–273.
Van Lieshout, M.N.M., Molchanov, I.S., 1998. Shot noise weighted processes: a new family of spatial point processes. Commun. Statist. Stoch. Model. 14 (3), 715–734.
Van Lieshout, M.N.M., 2011. A *J*-function for inhomogeneous point processes. Statist. Neerlandica 65 (2), 183–201.