ELEMENTA
Science of the Anthropocene

RESEARCH ARTICLE

Epidemiologically inspired approaches to land-use policy evaluation: The influence of the Rural Environmental Registry (CAR) on deforestation in the Brazilian Amazon

Marcelo A. Costa*, Raoni Rajão*, Marcelo C. C. Stabile[†], Andrea A. Azevedo[†] and Juliano Correa*

Environmental policy evaluation is crucial to determining if policy objectives were achieved. In most cases, some of the outcomes can be measured but a proper statistical analysis is difficult to achieve since the data may not represent a random sample (i.e., the data is biased), are not representative of the population or cannot be compared to a control group. This work adapts quasi-experimental statistical methods widely used in epidemiological studies that could be applied to land use policy evaluation in situations of relatively poor data. In order to test and develop this set of methods, we evaluated the effect of a land-use policy known as the rural environmental registry (CAR) on the reduction of deforestation rates in the Brazilian Amazon rainforest. The random variable of interest is the number of deforested hectares in given private properties and the statistic of interest is the difference of the annual deforestation rate between the properties before and after the policy intervention. Since no formal statistical distribution properly fit the data, non-parametrical approaches such as Monte Carlo simulations and Bootstrap were used. Data from the Brazilian states of Mato Grosso and Pará were used, with different time periods and three rural property size classes. Results show that the properties inside the Rural Environmental Registry have reduced their deforestation rates in some property classes and time periods, but this effect has not been systematic across time and space indicating that the policy is only partially effective. We conclude that the proposed statistical methods can be useful in environmental policy evaluation in different contexts due to low demands in terms of data availability and statistical distribution assumptions.

Keywords: land-use policy; spatial statistics; monte carlo bootstrap methods

1. Introduction

Environmental policy evaluation is one of the most important, yet often neglected, aspects of policy life-cycle (Crabbé and Leroy, 2008). Policy evaluation should provide two interrelated types of analysis: the measurement of impact and the development of a counterfactual scenario (Ferraro, 2009; Greenstone and Gayer, 2009; Ferraro and Miranda, 2014). In relation to the first type of analysis, to improve environmental governance practices it is crucial to determine whether a given outcome can be attributed to a driver, program, policy or intervention. In relation to the second type of analysis, a counterfactual describes something contrary to fact, used to reduce or eliminate cofounding biases from other variables. Following studies in the natural sciences, experimental research design has become standard in an increasing number of fields. It is based on the selection of a statistically significant sample of an underlying population and the separation of the sample units into two groups: a "treatment group" and a "control group". These two groups should be statistically similar in all respects, except for exposure to the treatment. The premise is that the randomized procedure reduces the bias, and the treatment outcomes can be compared to give a credible estimate of the effect of treatment (Campbell and Stanley, 1966; Shadish et al., 2001).

Experimental designs are difficult to apply outside of controlled laboratory environments, in situations in which it is not possible to obtain a random sample. This is the case with environmental policy evaluations in which unaffected areas, that are statistically similar, cannot be found. For this reason, approaches known as quasi-experimental designs have been created to measure the effect of given treatments. These are applied in contexts in which random selection and the strict separation of effects are beyond the researchers' control (Campbell and Stanley, 1966; Ferraro, 2009; Greenstone and Gayer,

* Laboratório de Gerenciamento de Serviços Ambientais, Departamento de Engenharia de Produção, Universidade Federal de Minas Geraís, Belo Horizonte, BR

† Instituto de Pesquisa Ambiental da Amazônia (IPAM), Belém, Pará, BR

Corresponding author: Marcelo A. Costa (macosta@ufmg.br)

2009). In land-use policy evaluation criteria, socio-economic indicators (Gibbs et al., 2015) such as spatial proximity (Nepstad et al., 2006) and temporal frames (Stewart-Oaten et al., 1986; Wiens and Parker, 1995; Smith, 2002) have been used to establish the similarity between areas and to evaluate differences between areas affected and unaffected by a given policy, a concept also known as before-after/control-intervention (BACI), where the design involves a control versus treatment group, and evaluates policy by comparing the situation of an area before and after the introduction of a given policy.

Quasi-experimental research designs using counterfactual analysis aim at evaluating potential outcomes under specific scenarios, or hypotheses, such as what was the most likely scenario if the area under study had not being under the influence of a given policy (Shadish et al., 2001; Ferraro, 2009). Thus, it would be possible to investigate potential environmental benefits of the policy (e.g. how much forest had been spared in a given period due to the creation of new protected areas). For example, cellular automata (Soares-Filho et al., 2006; Vega Orozco et al., 2012), econometric (Soares-Filho et al., 2010; Arima et al., 2014) and probabilistic bottom-up models (Rosa et al., 2013; Godar et al., 2014) have been used to establish counterfactual analysis and, consequently, to measure the overall effect of environmental policies in reducing deforestation and greenhouse gas emissions.

Despite the wide variety of quantitative methods available for the measurement of policy effects and counterfactual analysis, the literature does not present methods that are both general enough to be applied across different domains and sufficiently robust to provide reliable information about the effects of environmental policies. Furthermore, most of the methods mentioned above require a vast amount of economically and spatially explicit data that is often unavailable to policy-makers, especially in developing countries. In this context, this paper seeks inspiration from epidemiology to propose a set of robust yet data-light set of methods to evaluate the effect of land-use policies (Coulston and Riitters, 2003; Tonini et al., 2009; Fei, n.d.; Tuia et al., n.d.).

Statistical epidemiological models have been applied to associate the mean behavior of the number of cases of diseases to environmental or socio-economic variables or other relevant information (Jewell, 2003; Selvin, 2004). In these models, the dependent random variable can be defined as the number of cases observed among an underlying population at risk. An appropriate statistical model for this dependent random variable is the Binomial model. If the proportion of the cases with respect to the risk population is small then the Poisson distribution can be used as an approximation to the Binomial distribution. Furthermore, if further geographical information is available then spatial clustering analysis (Lawson, 2013) can be used to detect spatial clusters in which the disease rate is significantly higher. This information is crucial for early treatment of individuals and to stop the dissemination of contagious diseases.

We argue in this paper that statistical epidemiological models can also be useful for the analysis of environmental

policies such as the rural environmental registry (CAR) that is part of the Brazilian Forest Code. Created in 1965, the Forest Code (FC) was transformed during the 90's into the main Brazilian environmental federal law. The FC was revised in 2012, maintaining conservation requirements, including a Legal Reserve on at least 80% of native in private properties for the Amazon Biome. At the same time, the new FC provided an amnesty of all fines and of 58% of the areas illegally deforested in the past, while providing more flexibility for the compensation of the remaining areas with creation of the environmental reserve quota (CRA) market (Soares-Filho et al., 2014). In order to partially compensate for the amnesty and legitimization, the new FC built upon the experience of the states of Mato Grosso and Pará and created the national rural environmental registry, CAR (an acronym for Cadastro Ambiental Rural in Portuguese). CAR aims to document the degree of environmental compliance of more than 5 million rural properties in Brazil. Registry in the CAR is a voluntary initiative of the land owner, although mandatory under the FC. Benefits of joining the CAR include a lower chance of receiving fines for not complying with environmental laws, access to additional lines of credit for farmers, and the opportunity to sell to supply chains that have deforestation agreements, like those for soy and cattle (Azevedo et al., 2017).

From the perspective of epidemiology, deforestation can be conceptualized as a disease that affects an underlying population (i.e., forests), causing the decrease or death of the individuals (i.e., clearing of certain areas). Here the implementation of the CAR is assumed to be a treatment that could reduce the loss of individuals. Therefore, the proposed approach provides an example of quasi-experimental design with similarities to clinical trials and other epidemiological studies. The remainder of this paper is organized as follows: Section 2 presents the data sets for both Mato Grosso and Pará States; Section 3 presents the proposed statistical methods; Section 4 presents the results; and the discussion and conclusions are presented in Section 5.

2. The Rural Environmental Registry in Mato Grosso and Pará States

CAR is a registry implemented in the states of Mato Grosso and Pará in 2009 and 2008, respectively, with the aim to speed up the process of properties becoming compliant with Brazil's Forest Code and to improve the monitoring capabilities of the states. Thus, CAR is part of a land-use policy aiming to reduce illegal deforestation in the Brazilian Amazon. The registry contains georeferenced data of the borders, hydrography, and land-use of individual rural properties that can be combined with deforestation data provided by PRODES, a deforestation monitoring system developed by the Brazilian Institute for Space Research (INPE, http://www.obt.inpe.br/prodes/index.php). Currently, in these states there is no other dataset of rural properties as complete as CAR, and for this reason the location and land-use of these areas were not known prior to their entry in the registry. Similarly, it would be a mistake to compare the land-use dynamics of

the properties inside CAR, which tend to be active farms, with indiscriminate areas outside the registry that may include public undesignated lands and other areas that are not likely to be deforested in the near future. In order to deal with this challenge, a BACI quasi-experimental design was adopted to classify the properties, for every year of the analysis, into two groups. The properties that were already registered inside CAR in a given year were considered as part of the intervention group for the policy. Conversely, the properties that in a given year are not yet in the registry but that would join CAR in a future year are considered as part of the control group, based on the assumption that they are similar to the intervention group in all aspects except for not yet being under the influence of that policy. For instance, in 2010 for the state of Pará, the properties that joined CAR in 2008 and 2009 are part of the intervention group, whereas the control group comprises those properties that will join CAR from 2011 onwards.

To ensure the robustness of the statistical analysis, a substantial number of properties had to be excluded from the study. To account for the limitation in the spatial resolution of PRODES, which is unable to detect clearings under 6.25 hectares, all properties with areas of less than 10 ha were excluded from the dataset, as were properties outside of the Amazon Biome in Mato Grosso State. Properties with an accumulated deforestation greater than 95% per year were also excluded to eliminate the possibility of the deforestation rate being influenced by the absence of forest in a specific group of properties. Properties regularized under other land and environmental policies were also excluded to avoid possible interference in the analysis of CAR's effects on deforestation. Among the properties influenced by activities of land regularization and excluded from the study are rural settlement projects from the Brazilian Institute for Colonization and Agrarian Reform (INCRA, from the acronym for Instituto Nacional de Colonização e Reforma Agrária, in Portuguese) that are quite different from the properties that make up most of the private land in Pará and Mato Grosso. In the specific case of Mato Grosso State, properties which began the process of licensing (LAU, from the acronym for Licença Ambiental Única in Portuguese) before the creation of CAR's instrument were excluded, as this is a more encompassing policy that also

provides authorizations for legal deforestation. Finally, to improve the spatial consistency of the dataset, properties in CAR with more than 70% of their georeferenced area overlapped by neighboring properties were excluded because it was not possible to determine which of them were correct in the registry. In cases in which the overlap was smaller than 70%, manual inspection was used to exclude the property with the oldest CAR date. Furthermore, properties without a date of registration in CAR were excluded from our analysis. Therefore, 53.4% of the properties and 29.9% of the area from Pará State were not included in our analysis, leaving 19,466 CAR properties with a total area of 11.1 million hectares. In Mato Grosso State, 54.90% of the properties and 45.72% of the area from the original data were not included in our analysis, leaving 3,559 CAR properties with a total area of 3.1 million hectares. A more detailed description of the data for each state is given in the next section.

The dataset for both states were also subdivided to control for the effect of other factors in change of land-use. Specifically, to control for the effect of property size on deforestation rates, the dataset for both states were divided into three category groups according to their areas in terms of fiscal modules (FM), a measurement that varies for each municipality and is used as a criterion for the definition of legal rights and obligations. The first group consists of properties having up to four fiscal modules (one fiscal module represents 100 ha in most municipalities in the Amazon) that are considered "small properties" by law. In the second group are medium size properties that range between 4 and 15 FM (i.e., a property in the Amazon with 401 ha belongs to the medium size properties group). The third group are the large properties with more than 15 FM (i.e., usually more than 6,000 ha). Finally, in order to control for the effect of other regional policies (e.g. law enforcement actions, governmental subsidies) and economic factors (e.g. increases in commodity prices, variations in land price) that vary over time, the analyses were carried out with comparisons only within the same year.

2.1. Pará (PA) database
The sample used for Para State is described in **Table 1**. In Pará the number of small properties (up to 4 FM) is 13,487, 69.3% of the sample size and accounting for 11.42% of

Table 1: Descriptive Statistics for the Samples for Pará and Mato Grosso. DOI: https://doi.org/10.1525/elementa.260.t1

Property size group (FM)	Pará		Mato Grosso	
	Properties Sampled – qty	Area Sampled Thousands ha	Properties Sampled – qty	Area Sampled Thousands ha
up to 4 FM	13,487 (69.3%)	1,270 (11.42%)	1,923 (54.02%)	282 (9,01%)
from 4 to 15 FM	3,453 (17.7%)	2,040 (18.32%)	1,041 (29.26%)	739 (23.62%)
over 15 FM	2,527 (12.98%)	7,810 (70.27%)	595 (16.72%)	2,109 (67.37%)
TOTAL	19,467	11,120 ha	3,559	3,130 ha (t)

Table 2: Dynamics of the enrollment of properties in CAR, from 2008 to 2012 in the state of Pará. DOI: https://doi.org/10.1525/elementa.260.t2

Property size group (FM)	Enrollment of the properties	2008	2009	2010	2011	2012
up to 4 FM	Before CAR (control)	99.40%	96.80%	75.44%	45.08%	3.48%
	CAR	0.60%	3.20%	24.56%	54.92%	96.52%
from 4 a 15 FM	Before CAR (control)	99.25%	90.90%	59.58%	34.67%	3.49%
	CAR	0.75%	9.10%	40.42%	65.33%	96.51%
greater than 15 FM	Before CAR (control)	97.66%	87.57%	52.13%	31.55%	2.55%
	CAR	2.34%	12.43%	47.87%	68.45%	97.45%

Table 3: Dynamics of the enrollment of properties in CAR, from 2009 to 2011 in the state of Mato Grosso. DOI: https://doi.org/10.1525/elementa.260.t3

Property size group (FM)	Enrollment of the properties	2009	2010	2011
up to 4 FM	not CAR (control)	97.27%	41.03%	1.10%
	CAR	2.73%	58.97%	98.90%
from 4 a 15 FM	not CAR (control)	90.36%	35.74%	1.66%
	CAR	9.64%	64.26%	98.34%
over 15 FM	not CAR (control)	89.60%	40.21%	2.08%
	CAR	10.40%	59.79%	97.92%

the sample area. Medium properties total 3,453 (17.7%), while occupying 18.32% of the sample area. There are 2,527 large properties (12.98%), with 70.27% of the sample area. In terms of area, small properties accounted for 1.27 million ha, the medium properties 2.04 million ha, and large properties 7.81 million ha. Therefore, the group of properties with more than 15 FM represented 70.27% all areas inside the dataset.

The time dynamics of enrollment of properties in CAR is shown in **Table 2**. In general, in 2008, less than 2.54% of the properties in the different size groups had been enrolled into CAR. By 2012, more than 96.5% of the properties had been enrolled in CAR.

2.2. Mato Grosso (MT) database

The sample used for Mato Grosso State is described in **Table 1**. The database of the state of Mato Grosso has 3,559 properties divided into three groups. The first group consists of properties of up to four fiscal modules with 1,923 (54.02%) properties and 9.01% of the sample area. The second group consists of properties with 4 to 15 modules with 1,041 properties (29.26%) and 23.62% of the area. The third group consists of properties with more than 15 modules and has 595 (16.72%) properties, accounting for 67.37% of the sample area. Properties with up to four FM had 282,006 ha; the second group had 739,756 ha; and the third group had 2,109,579 ha. Therefore, the group of properties with more than 15 FM represented 67.37% of the total sample area.

The time dynamics of the enrollment of the properties in the CAR is shown in **Table 3**. In general, in 2009, fewer than 11% of the properties in the different size groups had not been enrolled into CAR. By 2011, more than 97% of the properties had been enrolled in the CAR.

3. Statistical methods

Let Y_{it} be a random variable representing the deforested area (in ha) of property i at time t. F_{it} represents the area of forest (in ha) of property i at time t. Each property can be classified into one of three groups related to size. Let $j_{[i]}$, $j_{[i]} \in \{1, 2, 3\}$ be the index related to each size group, or simply j. For properties up to 4 FM, then $j = 1$; for properties from 4 to 15 FM, then $j = 2$; and for properties greater than 15 FM, then $j = 3$. Furthermore, each property i can be classified through time t into the CAR or Control groups. Thus, let $k_{[i,t]}$ be the index representing the CAR ($k = 1$) or Control ($k = 2$) groups) of property i at time t.

Assume that Y_{it} is a random variable that represents the number of hectares deforested in property i at time t. Let F_{it} be total number of hectares of forest in property i at time t. If Y_{it} were a discrete random variable, then one may identify the potential distribution of the random variable Y_{it} as a Binomial distribution:

$$Y_{it} \sim Binomial\left(Y_{it} + F_{it}, \rho_{it}\right) \qquad (1)$$

where ρ_{it} represents the probability of deforestation of one ha in property i at time t, $P(Y_{it} = 1) = \rho_{it}$, and $Y_{it} + F_{it}$ is the total number of forest hectares at the beginning of time t in property i. It is known that maximum likelihood estimates require the probability distribution of the observations. However, quasi-likelihood estimates

(Wedderburn, 1974) requires only the relation between the mean and the variance of the observations. Thus, assuming that Y_{it} is a continuous random variable with mean $E(Y_{it}) = (Y_{it} + F_{it}) \cdot \rho_{it}$ and variance of $Var(Y_{it}) = (Y_{it} + F_{it}) \cdot \rho_{it} \cdot (1 - \rho_{it})$ then Equation 1 is also a possible approximation to the stochastic behavior of the random variable Y_{it}. However, we found that in practice, the empirical distribution of the random variable did not fit the Binomial distribution, i.e., the deviance quality-of-fit statistic (Nelder and Baker, 1972) indicated the rejection of the null hypothesis of model fit (P-value = 0.0000). Alternatives such as Poisson, Negative Binomial, Zero Inflated Poisson and Quasi-Likelihood models were also evaluated but did not fit the deforestation data.

As an alternative solution, we evaluated the first moment, i.e., the mean of the random variable, hereafter defined as $E(Y_{it}) = \tau_{jkt}$. Confidence intervals and prediction intervals were generated using Monte Carlo simulations and Bootstrap resampling techniques.

3.1. Statistical comparison of the effect of CAR using Monte Carlo simulations

Monte Carlo simulations (Dwass, 1957) are widely applied to provide statistical inference when the underlying distribution of the statistic of interest, or test statistic, is unknown. In many cases, the probability distribution of the test statistic given a null hypothesis cannot be calculated. Nevertheless, samples from the null distribution can be drawn using simulations. For example, in spatial disease clustering analysis (Kulldorff, 1997), the statistical inference of a cluster candidate, given the null hypothesis of spatial randomness, is conducted by randomly assigning cases to areas in proportion to their underlying at-risk populations. In sequence, the test statistic of the most likely cluster in the simulated scenario is stored. The procedure is repeated many times, say 9,999 and the p-value is the proportion of simulations in which the simulated test statistic was higher (or lower) than the observed test statistic, using the original data set. Glasserman (Glasserman, 2003), for instance, applies Monte Carlo simulations in economic settings to evaluate the performance of test statistics under simulated economic scenarios. As previously mentioned, our random variable of interest is the deforested area. We rely on statistical methods from spatial epidemiology (Lawson, 2013), since deforested areas behave similarly to observed disease cases from underlying populations, which are forests. That is, we assume that deforested areas are observed cases from an underlying continuous population of forested areas.

For each size group j and class k, the annual deforestation rates are calculated using Equation 2:

$$\tau_{jkt} = \frac{\sum_{j_{[l]}k_{[i,t]}} Y_{it}}{\sum_{j_{[l]}k_{[i,t]}} (F_{it} + Y_{it})} \quad (2)$$

It is of interest to test whether the deforestation rates within the different size groups ($k = \{1, 2\}$), at time t are

similar. For this question, the following null hypothesis can be written:

$$H_0 : \tau_{j1t} = \tau_{j2t} \quad (3)$$

In practice, we want to test the null hypothesis that the deforestation rates between properties that have adopted the CAR over the years are similar to the deforestation rates of properties that have not adopted CAR over the years. For this question, we use a Monte Carlo simulation procedure to perform hypothesis testing. The algorithm is as follows:

Monte Carlo simulation procedure to test the null hypothesis of similarity between deforestation rates among properties which adopted CAR and the properties that did not adopt CAR.

1. For each size group and time, calculate the deforestation rates of properties that adopt CAR and properties that did not adopted CAR:

$$\tau_{j1t} = \frac{\sum_{j_{[l]},1} Y_{it}}{\sum_{i \in j_{[l]},1} (F_{it} + Y_{it})} \quad \tau_{j2t} = \frac{\sum_{j_{[l]},2} Y_{it}}{\sum_{i \in j_{[l]},2} (F_{it} + Y_{it})} \quad (4)$$

2. Let the test statistic be the difference between the deforestation rates of properties that adopt CAR and properties that did not adopted CAR:

$$\kappa_{jt} = \tau_{j1t} - \tau_{j2t} \quad (5)$$

3. Conditional on the number of properties with non-zero deforestation at time t and size group $j(n_{jt})$, and the total number of deforestation at time t and size group, the distribution of the test statistic under the null assumes that the number of deforested units (in hectares) in randomly selected n_{jt} properties are proportional to the total number of forested areas in these properties:
   a. n_{jt} properties are randomly selected with no replacement.
   b. Let Y_t be the total number of deforested areas at time t, $Y_t = \Sigma_i Y_{it}$. Let $p_{i^*t} = \frac{Y_{i^*t} + F_{i^*t}}{Y_t + F_t}$, $i^* \in n_{jt}$ and $F_t = \Sigma_{i^*} F_{i^*t}$ be the proportion of deforested and forested areas in property i^*. In this case, under the null, it is possible to simulate the deforestation for each property i^* at time t using a multinomial distribution.

$$Y_{it} \sim multinomial(Y_t, \pi = [p_{1t}, \ldots, p_{it}]) \quad (6)$$

4. S simulations are generated using the multinomial distribution. For each simulation, the test statistic (Equation 5) is calculated. Thus, a sample of the test statistic under the null hypothesis is obtained, $\left(\kappa_{jt}^{(1)}, \kappa_{jt}^{(2)}, \ldots, \kappa_{jt}^{(S)}\right)$.

5. Finally, the values of $\kappa_{jt}^{(\cdot)}$ are ordered and if the observed statistic is less than the 2.5% percentile or greater than the 97.5% percentile of the simulated values, then the null hypothesis is rejected at the 5% level.

6. In addition, P-value estimates are based on the rank of the observed statistic with respect to the simulated values.

3.2. Counterfactual of CAR policy using bootstrap resampling for statistical forecasting

Following the estimation of the annual deforestation rates for each size group, it is of interest to compare the observed values of the remaining forest areas with a hypothetical scenario in which all properties did not adopt CAR. In this case, we aim at providing further evidence of the effects of CAR policy in reducing the deforestation and, therefore, resulting in larger areas of remaining forest. To account for the estimates and the associated variability we propose a bootstrap resampling procedure.

The bootstrap resampling procedure (Efron, 1979) is similar to Monte Carlo simulations. However, Monte Carlo simulations generally require the specification of a null hypothesis or a scenario from which samples are drawn. The Bootstrap procedure aims at estimating the distribution function F which generated the observed random sample, $Y_1, Y_2, ..., Y_n$. To do so, it creates bootstrap samples, $Y_1^*, Y_2^*, ..., Y_n^*$, which are random samples from the original data set with replacement. These bootstrap samples can be used to estimate confidence intervals of a test statistic of interest. For example, suppose we want to estimate a confidence interval for the sample mean, $\bar{Y}$. Thus, we can generate B bootstrap samples of size n and, for each bootstrap sample, calculate the bootstrap sample mean, $\bar{Y}_1^*, \bar{Y}_1^*, ..., \bar{Y}_B^*$. Confidence intervals are provided using the rank of the bootstrap sample means. Further details about bootstrap estimates can be found in Efron and Tibshirani (Efron and Tibshirani, 1994) and Dekking (Dekking, 2005).

In our case, we want to get samples from the distribution of deforestation rates for properties which did not adopt CAR and use these samples to forecast the behavior of all properties, if they had never been enrolled in CAR. We do this to provide further statistical evidence regarding the effectiveness of the CAR policy. Our proposed bootstrap procedure is shown next.

Bootstrap procedure to evaluate the hypothetical scenario in which all properties had never been enrolled in CAR:

1. First, the period of interest is held fixed: $t_0, t_1, ..., t_f$. For the state of Pará the period the simulation is from 2007 to 2012, and for the state of Mato Grosso period of simulation is from 2008 to 2011. Thus, t_0 is the baseline. The total forest area for the baseline was set as the reference level (100%).

2. From t_1 to t_f the deforestation values, and consequently, the remaining forest were estimated for all properties but assuming they had the deforestation rates of the properties which did not adopt the CAR. Thus, the forecast of remaining forest areas over the period of interest (t_1 to t_f), and for the different size groups is given by equation 7:

$$F_{j,t+1} = F_{j,t} \times \left(1 - \tau_{j2t}\right) \tag{7}$$

To account for uncertainties related to the estimated deforestation rates and, consequently, estimate the uncertainties in the forecasted remaining forests, the following bootstrap resampling procedure was used:

3. Using the observed values of forest and deforested areas in the database (Y_{it}, F_{it}), B bootstrap samples were generated ($b = 1, ..., B$). Each replicate has the same size of the original database, and it has been generated using resampling from the original database with replacement. For each bootstrap sample, the rates of deforestation were calculated using Equation 8, but only for $k = 2$ (properties that did not adopt CAR):

$$\tau_{jkt}^b = \frac{\sum_{j_{[i]}k_{[i,t]}} Y_{it}^b}{\sum_{j_{[i]}k_{[i,t]}} \left(F_{it}^b + Y_{it}^b\right)} \tag{8}$$

4. The bootstrap deforestation rates, calculated using Equation 8, were used to forecast the deforestation and forest areas for all properties as if they had not adopted CAR in the period, starting from the original baseline forest. Equation 9 shows the forecast equation:

$$F_{j,t+1} = F_{j,t} \times \left(1 - \tau_{j2t}^b\right) \tag{9}$$

5. The 2.5 and 97.5 percentiles of the B bootstrap forecasts were used to create an empirical bootstrap interval with 95% confidence. These intervals represent the projection of remaining forest areas throughout the studied period, assuming a scenario in which no properties had adopted CAR.

3.3. The space-time scan statistic

The space-time scan statistic (Kulldorff et al., 1998) aims at detecting clusters in space and time in which the observed number of cases is significantly higher than the expected number of cases, under the null hypothesis of space-time randomness. It scans the 3-dimensional space defined by the spatial geographical coordinates and the time period using a cylindrical window, as shown in **Figure 1**. The base of the cylinder represents the spatial component whereas the height represents the time range. Both the center of the base and the height of the cylinder are varied. By changing the location of the base, its radius, the starting and stopping times (i.e., the height) of the cylinder, different configurations are created. For each configuration, the observations inside and outside the cylinder are used to calculate a likelihood ratio statistic. The base and height configuration, i.e., the cylinder configuration with the maximum value of the likelihood ratio function (see Equations 10 and 11) represents the final cluster candidate. Secondary clusters can also be evaluated by selecting non-overlapping cylinders with large likelihood ratio statistics. Statistical inference is performed using Monte Carlo simulations, which provide statistical evidence for accepting or rejecting the null hypothesis. Further details are shown below.
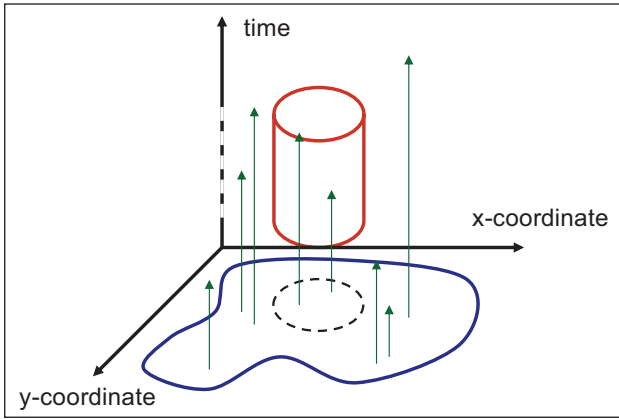
Figure 1: Space-time scan statistic using a cylinder scanning windows with variations in the height (time) and location (spatial component). The space-time scan statistic that aims at detecting clusters in space and time in which the observed number of cases is significantly higher than the expected number of cases, under the null hypothesis of space-time randomness. DOI: https://doi.org/10.1525/elementa.260.f1

The space-time scan statistic is a widely-used method to detect clusters in epidemiological settings. In our case, we consider the Forest units as the at risk population and the deforested units as disease cases. Under the null hypothesis of space-time randomness, the deforested units are uniformly distributed in the population (Forest units). Therefore, the number of deforested units in property i at time t is Poisson distributed with the expected number of cases, μ_{it}, proportional to the size of Forest units in the previous year ($F_{i,t-1}$):

$$H_0 : Y_{it} \sim Poisson\left(\mu_{it} = \lambda F_{i,t-1}\right) \qquad (10)$$

where $F_{i,t-1} = F_{it} + Y_{it}$. Under the null hypothesis, $\hat{\lambda} = D/F$, where D is the total number of cases (deforested units) in space and time, and F is the total population (forest units) in space and time. Under the alternative hypothesis, there is one space-time cluster at an unknown location. Define Z as the set of all possible cylinder clusters z. For each cluster z ($z \in Z$) let d_z and F_z be the number of deforested units and Forest units inside cluster z. The likelihood ratio test statistic associated with the most likely cluster is written as:

$$\frac{L\left(\hat{z},\hat{p},\hat{r}\right)}{L_0} = sup\left(\frac{d_z}{\mu_z}\right)^{d_z}\left(\frac{D-d_z}{D-\mu_z}\right)^{D-d_z} \qquad (11)$$

Where μ_z is the expected number of cases under the null hypothesis, $\mu_z = D \cdot d_z/F$. Monte Carlo simulations (Dwass, 1957) are applied to address the statistical significance of the most likely cluster. Further details can be found in Kulldorff et al. (Kulldorff et al., 1998) and Costa and Kulldorff (Costa and Kulldorff, 2014). The R code used in all our calculations is given as Supplemental information (Data S1).

4. Results
In epidemiological settings, the number of cases of diseases can be modeled as Poisson distributed with the expected number of cases proportional to the at-risk population. Our data set has a large number of properties with zero deforested areas. For the state of Mato Grosso, 95.7% of the records have zero deforested areas. For the state of Pará 84.1% of the records have zero deforested areas. Parametric statistical models, assuming a Poisson statistical distribution or a Negative Binomial, did not fit the data., i.e., the deviance quality-of-fit statistic indicated the rejection of the null hypothesis of model fit (P-value = 0.0000). Therefore, our final analysis is based on non-parametric modeling, using Monte Carlo simulations and Bootstrap, as described in section 3.

Table 4 shows the estimated deforestation rates for each 100 ha, for each property size group and year. The Monte Carlo inference results are shown in the last column. P-values greater than 5% (0.05) indicate that the null hypothesis that the deforestation rates without CAR and with CAR are similar, was not rejected. Results show that in 2010, for properties in the state of Mato Grosso with size up to 4 FM, there is evidence that the deforestation rates between CAR and not CAR properties are different. This is also true for properties of size from 4 to 15 FM and over 15 FM in the state of Mato Grosso, in 2009. For Pará, there is statistical evidence that the deforestation rates between CAR and not CAR are different for properties of size up to 4 FM for all years, except in 2012. This is also true for properties of size from 4 to 15 FM, except in 2009. For properties of size over 15 FM there is statistical evidence that the deforestations rates are different only in 2012.

Figure 2 shows the deforestation rates for each size group from 2009 to 2011 in the state of Mato Grosso. For properties with sizes of up to 4 FM, the deforestation rates increase in the period for both CAR and without CAR properties. For properties with sizes from 4 to 15 FM, the deforestation rates have higher values in 2010 and a slight decrease in 2011 for both CAR and without CAR properties. For properties with sizes over 15 FM, the deforestation rates present a slight decrease for properties without CAR, whereas for properties with CAR the deforestation rates present a slight increase.

Figure 3 shows the deforestation rates for each size group from 2008 to 2012 in the state of Pará. In general, the deforestation rates decrease in time for both properties which have enrolled in CAR and for those which have not enrolled in CAR. Properties which have enrolled in CAR present deforestation rates smaller than those properties which have not enrolled in CAR.

Figures 2 and **3** shows that, in general, properties with sizes over 15 FM presented smaller deforestation rates compared to properties with sizes of up to 4MF, and properties from 4 to 15 FM.

Table 5 presents the total forest area for the different size groups, at the baseline period which is 2007 for the state of Pará and 2008 for the state of Mato Grosso. These values are the sum of deforested and forest areas at the baseline. They were used as the reference

Table 4: Comparative analysis of estimated deforestation rates between properties which adopted CAR and those which did not adopt CAR for the states of Pará and Mato Grosso. DOI: https://doi.org/10.1525/elementa.260.t4

Property size group	Year	Estimated deforestation rates for each 100 ha		CAR effect in deforestation rates (CAR/control − 1)	P-value H_0: without CAR = with CAR
		Control (without CAR)	CAR (with CAR)		
State of Mato Grosso (2009–2011)					
up to 4 FM	2009	0.9444	0.5672	−39.9%	**0.4688**
	2010	1.4685	0.5967	−59.4%	0.0137
	2011	1.6609	1.7424	4.9%	**0.5796**
from 4 to 15 FM	2009	0.2349	0.6956	196.1%	0.0162
	2010	0.9089	0.7476	−17.7%	**0.3016**
	2011	0.4001	0.5150	28.7%	**0.5675**
over 15 FM	2009	0.0988	0.0046	−95.3%	0.0258
	2010	0.0818	0.0954	16.5%	**0.3980**
	2011	0.0286	0.2201	669.9%	**0.3694**
Pará (2008–2012)					
up to 4 FM	2008	4.5769	0.9037	−80.26%	0.0002
	2009	3.5037	0.5933	−83.07%	0.0001
	2010	3.7155	2.4231	−34.78%	0.0001
	2011	2.6247	2.2073	−15.90%	0.0071
	2012	1.8443	1.6189	−12.22%	**0.2600**
from 4 to 15 FM	2008	2.9747	0.0196	−99.34%	0.0003
	2009	2.0516	1.8890	−7.93%	**0.3109**
	2010	1.4162	0.8856	−37.47%	0.0001
	2011	0.7118	0.5689	−20.08%	0.0408
	2012	0.7658	0.4002	−47.74%	0.0404
over 15 FM	2008	1.2406	1.8518	49.26%	**0.1394**
	2009	0.5014	0.6694	33.50%	**0.2722**
	2010	0.4425	0.5759	30.14%	**0.4423**
	2011	0.1604	0.2239	39.58%	**0.5452**
	2012	0.0227	0.2144	846.17%	0.0876

values (100%) in the bootstrap simulations. **Table 6** compares the observed deforested and forest areas with the simulated forest areas as if the properties have not enrolled in CAR policy. In addition, upper and lower limits with 95% confidence are provided. These results are also presented in **Figures 4** and **5**. **Figure 4** compares the observed forest areas with the simulated forest areas if the properties had not enrolled into CAR policy in the state of Mato Grosso. The observed values for properties of the size groups "up to 4 FM" and "from 4 to 15 FM" are within the lower and upper limits which suggests that the observed levels of forest conservation could have happened even if no property

had adopted the CAR policy. For properties with sizes over 15 FM the simulation results suggest that if the properties had not adopted CAR policy then the remaining forest areas in 2011 would have been larger than the observed.

Figure 5 compares the observed forest areas with the simulated forest areas if the properties had not enrolled into CAR policy in the state of Pará. As opposed to what was observed for the state of Mato Grosso, for properties of the size groups "up to 4 FM" and "from 4 to 15 FM" the observed remaining forest areas at the end of the period are above the upper limit of the simulated confidence interval. This, suggests that the CAR policy succeeded
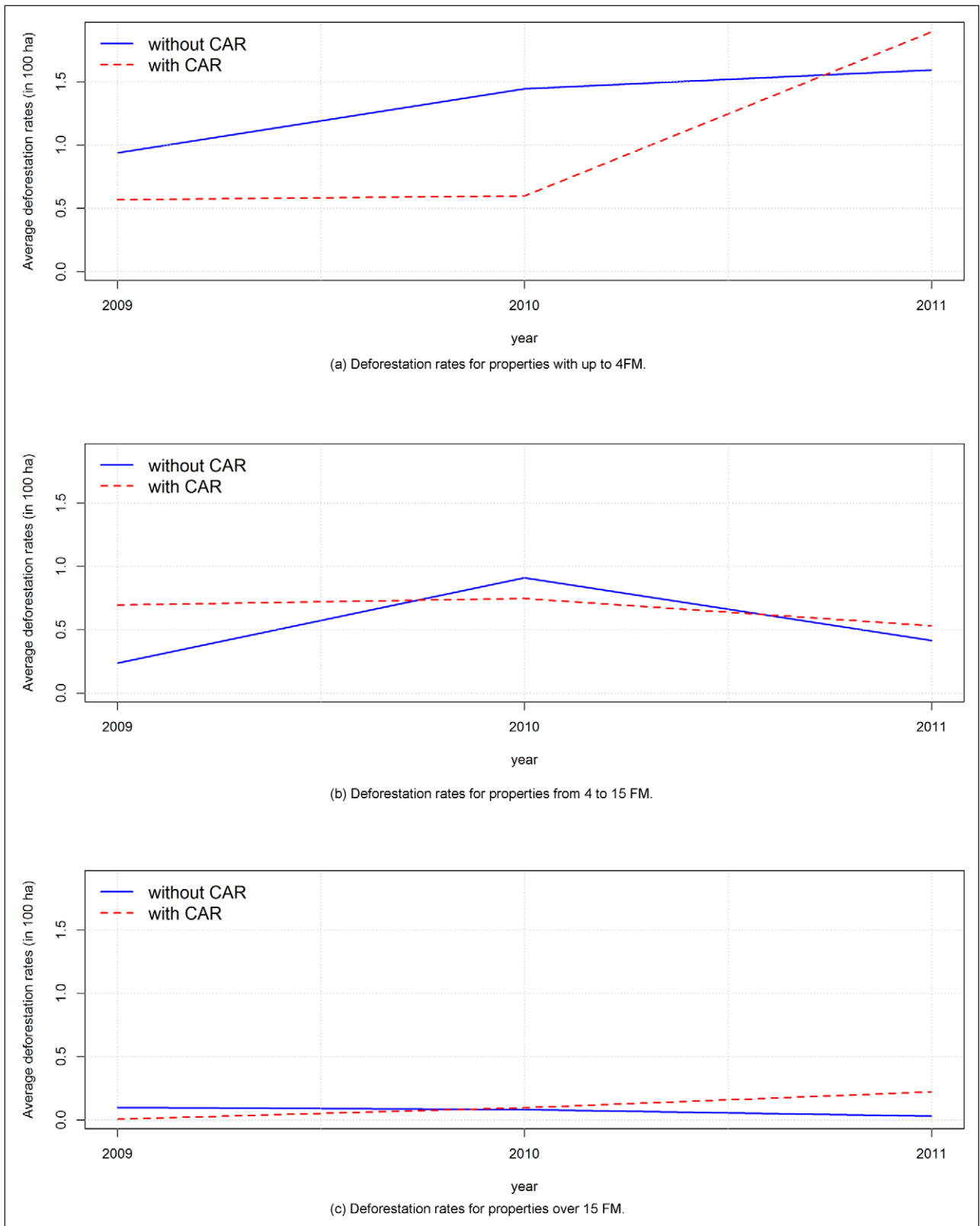
(a) Deforestation rates for properties with up to 4FM.

(b) Deforestation rates for properties from 4 to 15 FM.

(c) Deforestation rates for properties over 15 FM.

Figure 2: Mato Grosso Deforestation. Estimated deforestation rates for the state of Mato Grosso from 2009 to 2011. DOI: https://doi.org/10.1525/elementa.260.f2

in reducing deforestation rates and, consequently, resulting in larger forest area as compared to the scenario in which none of the properties had adopted CAR. It is worth mentioning that even though there was statistical evidence of the effectiveness of the CAR policy, there was no increase whatsoever in the forest area. Similarly to the case with Mato Grosso State, for properties with sizes over 15 FM the simulation results suggest that

(a) Deforestation rates for properties with up to 4FM.

(b) Deforestation rates for properties from 4 to 15 FM.

(c) Deforestation rates for properties over 15 FM.
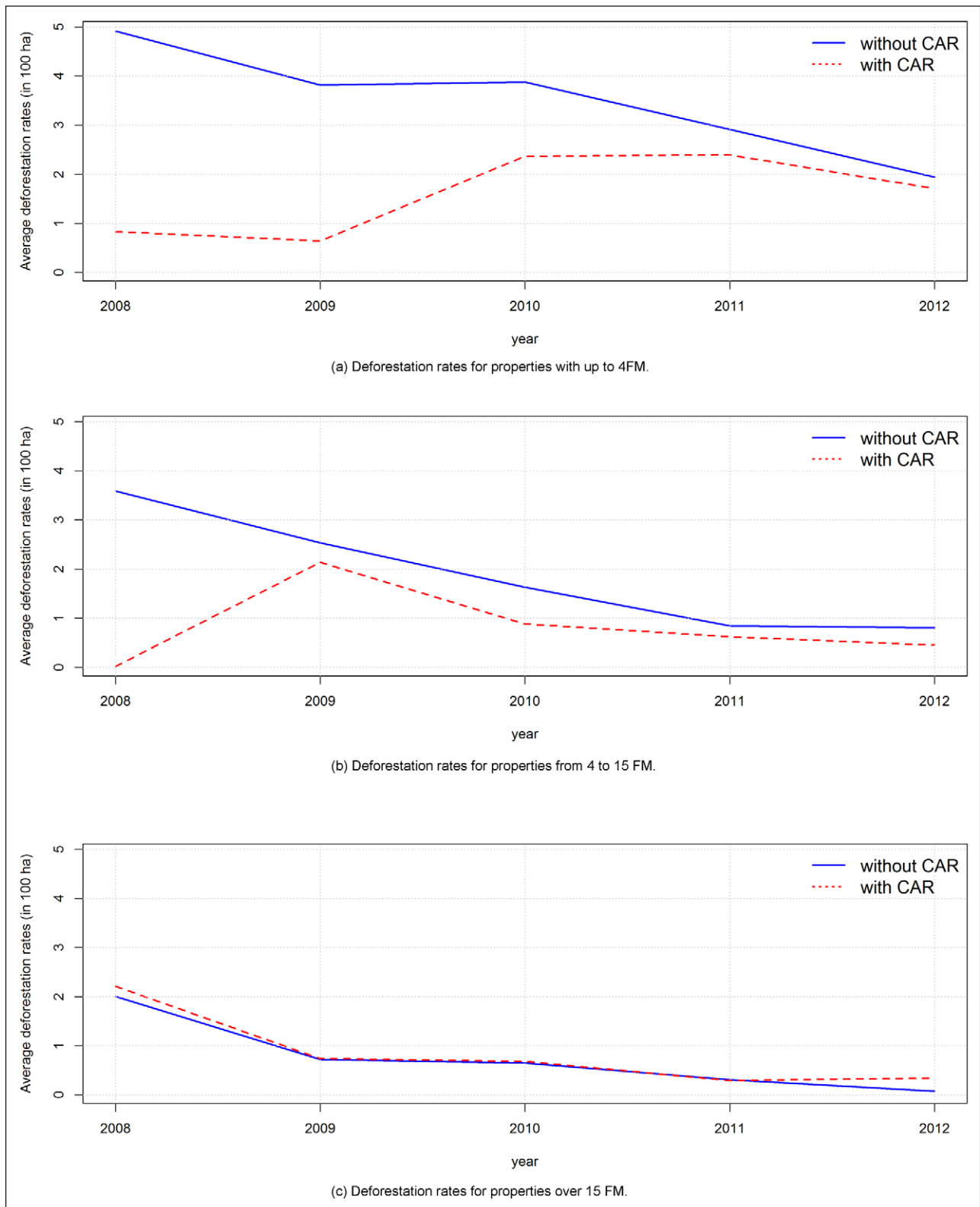
Figure 3: Pará Deforestation. Estimated deforestation rates for the state of Pará from 2008 to 2012. DOI: https://doi.org/10.1525/elementa.260.f3

if the properties had not adopted CAR policy then the remaining forest areas in 2012 would have been larger than the observed.

Figures 6 and **7** show the space-time cluster analysis. For visual representation of the results, a maximum cluster size parameter of 50% of the records for the state of Mato Grosso and 20% of the records for the state of Pará were chosen. Originally, a maximum cluster size parameter of 50% were applied to both states, as suggested by and Costa (Ribeiro and Costa, 2012). However, only one large

Table 5: Forest areas (in ha) for the different size groups, used as baseline in the Bootstrap deforestation simulation algorithm. DOI: https://doi.org/10.1525/elementa.260.t5

State	Baseline year	Properties size group		
		Up to 4 FM	from 4 to 15 FM	over 15 FM
Pará	2007	616.398	1.058.964	3.947.784
Mato Grosso	2008	101.474	359.345	1.253.688

Table 6: Forest areas (in ha) for the different size groups and simulated forest area as if the properties have not enrolled in CAR policy. DOI: https://doi.org/10.1525/elementa.260.t6

Property size group	Year	Deforestation area	Forest area	Simulated forest area (without CAR)	Lower limit (2,5%)	Upper limit (97,5%)
Mato Grosso (2009–2011)						
Up to 4 FM	2009	947	100527	100515.7	100129.5	100859.7
	2010	1000	99527	99039.6	97976.8	99739.8
	2011	1733	97794	97394.6	93808.5	99039.6
from 4 to 15 FM	2009	960	358385	358500.9	357734.5	358994.7
	2010	2902	355483	355242.4	350312.6	357990.9
	2011	1820	353663	353821.0	350128.9	355242.4
over 15 FM	2009	1157	1252531	1252449.9	1251226.5	1253175.9
	2010	1123	1251408	1251425.0	1250405.6	1252090.3
	2011	2708	1248700	1251067.2	1250126.9	1251425.0
Pará (2008–2012)						
Up to 4 FM	2008	28045	588353	588186.3	586690.1	589628.2
	2009	19991	568362	567578.2	566071.0	569137.5
	2010	19176	549186	546489.9	545100.6	547913.5
	2011	13131	536055	532146.2	530709.3	533499.0
	2012	8717	527338	522331.9	518833.3	525798.6
from 4 to 15 FM	2008	31350	1027614	1027463.4	1024017.9	1030687.6
	2009	20974	1006640	1006383.8	1002163.0	1010135.6
	2010	12374	994266	992131.2	989055.2	994627.0
	2011	6208	988058	985069.0	983109.5	986736.1
	2012	4092	983966	977525.5	972548.1	981816.0
over 15 FM	2008	49553	3898231	3898806.6	3886366.3	3908580.8
	2009	20285	3877946	3879258.0	3872354.9	3884091.5
	2010	19050	3858896	3862092.8	3856977.6	3865488.7
	2011	7452	3851444	3855897.5	3851212.5	3857985.7
	2012	7850	3843594	3855023.9	3851823.9	3855755.0

cluster was detected in the state of Pará. In this case, a smaller maximum cluster of 20% was chosen in order to improve cluster detection (Ribeiro and Costa, 2012). Detected clusters with a P-value smaller than 5% (0.05) are shown in **Figures 6** and **7**. The Poisson model was applied and, as previously mentioned, the Poisson distribution did not fit the data. Therefore, it is believed that the P-values are overestimated. Nevertheless, the results do provide important insights about areas which were more vulnerable to deforestation.
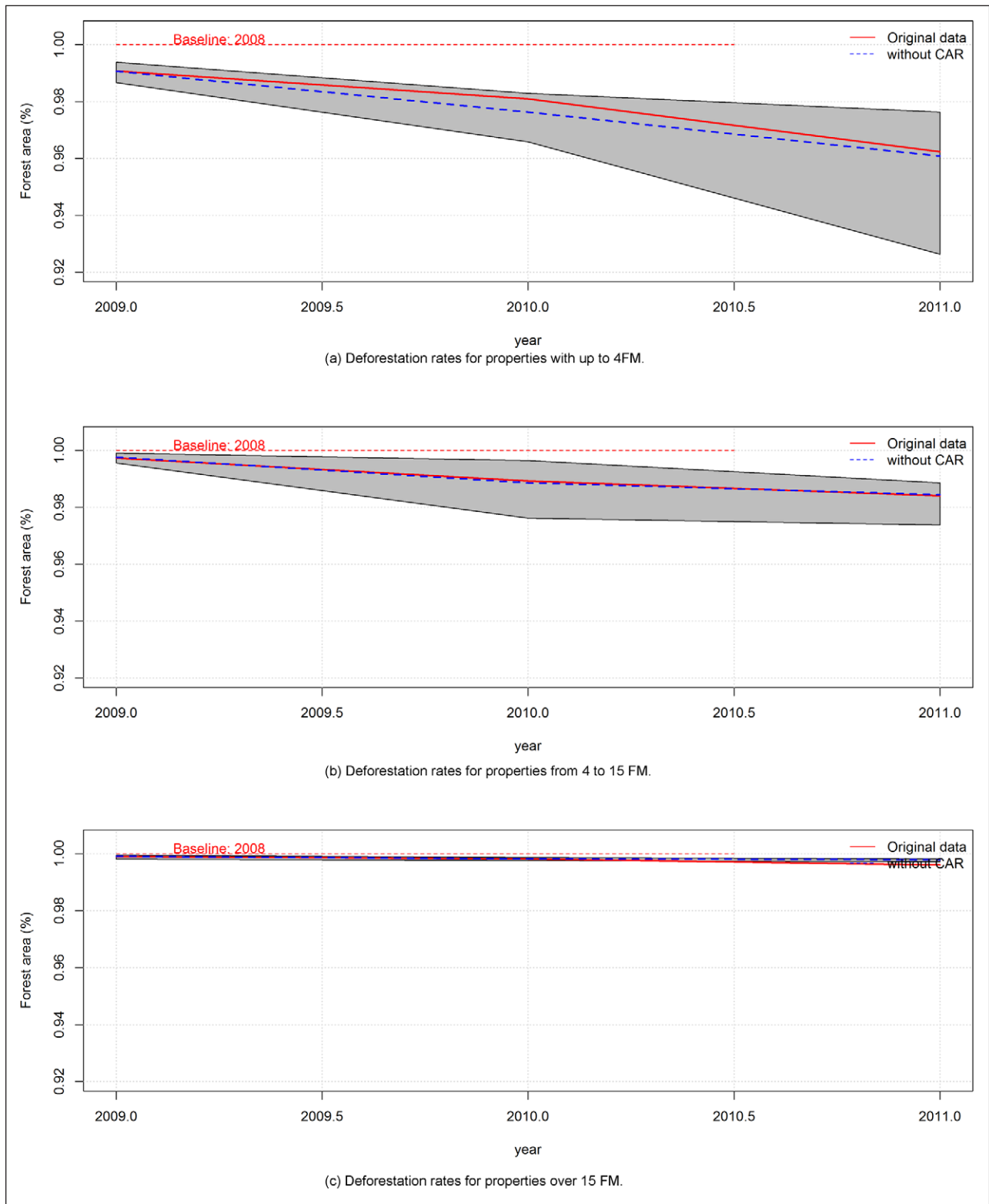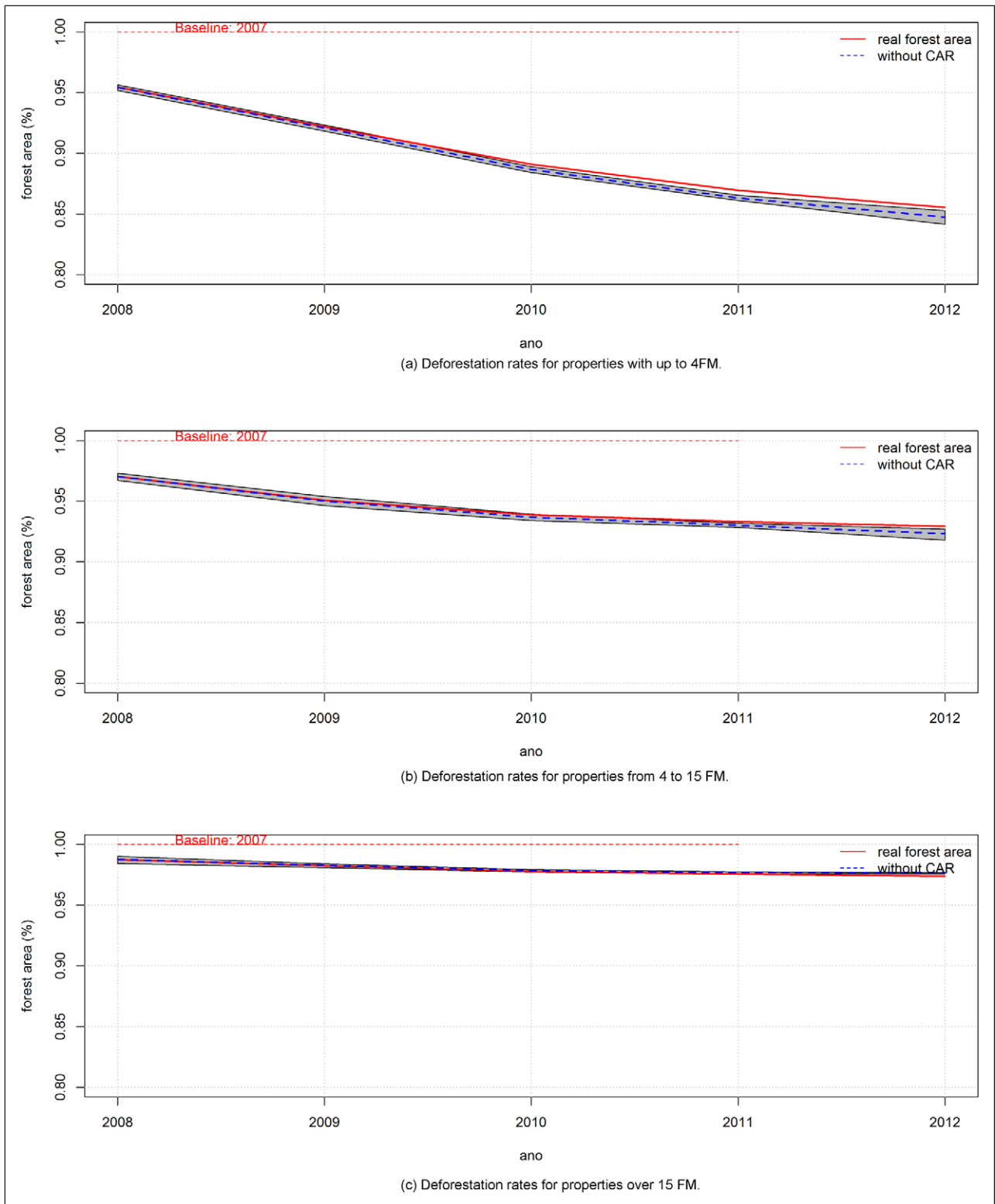
Figure 4: Simulation of remaining forest in Mato Grosso if properties did not adopt CAR. Simulation results for remaining forest areas in the state of Mato Grosso assuming the hypothetical scenario in which properties did not adopt CAR. DOI: https://doi.org/10.1525/elementa.260.f4

Figure 6(a) shows that detected clusters in the state of Mato Grosso are quite small. Squares represent clusters which include only one property. Therefore, there are few areas in which the deforestation rates are higher than expected under the null hypothesis of space-time randomness. **Figure 6(b)** shows that the detected clusters numbers 2, 4, 5, 6, 7 and 8 were found to be statistically significant in 2011. Clusters 1, 8 and 19 were statistically significant in 2010, and the remaining detected clusters were found to be statistically significant
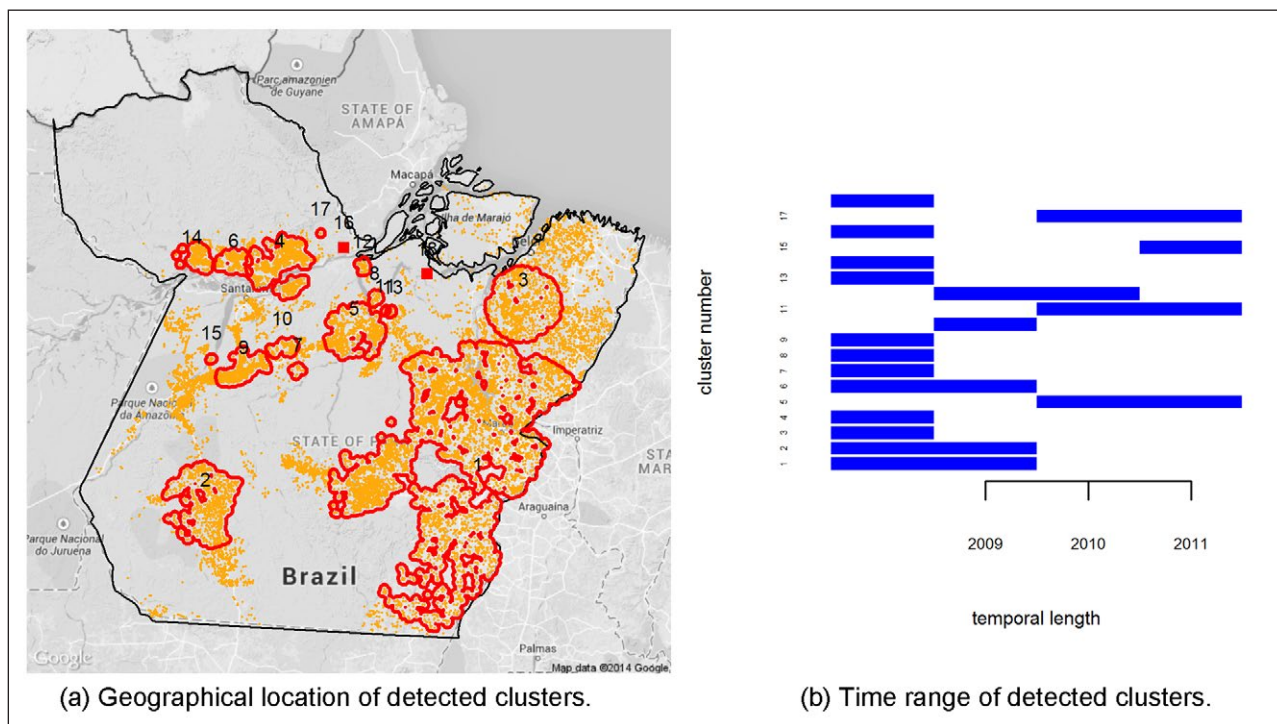
(a) Deforestation rates for properties with up to 4FM.

(b) Deforestation rates for properties from 4 to 15 FM.

(c) Deforestation rates for properties over 15 FM.

Figure 5: Simulation of remaining forest in Pará if properties did not adopt CAR. Simulation results for remaining forest areas in the state of Pará assuming the hypothetical scenario in which properties did not adopt CAR. DOI: https://doi.org/10.1525/elementa.260.f5

in 2009. Therefore, among the detected clusters, most of them happened in 2009 and 2010.

Figure 7(a) shows that the detected clusters in the state of Pará are larger as compared to the detect clusters in the state of Mato Grosso. Cluster one has a size which is more than 20% of the state area. This cluster was detected in years 2008 and 2009. Clusters 5, 11, 15 and 17 were active in 2011, which is the last year of data. As compared to the state of Mato Grosso, the state of Pará presented larger cluster areas. The state of Pará also presented larger deforestation rates in the clusters, as compared to its global rate.

(a) Geographical location of detected clusters.

(b) Time range of detected clusters.
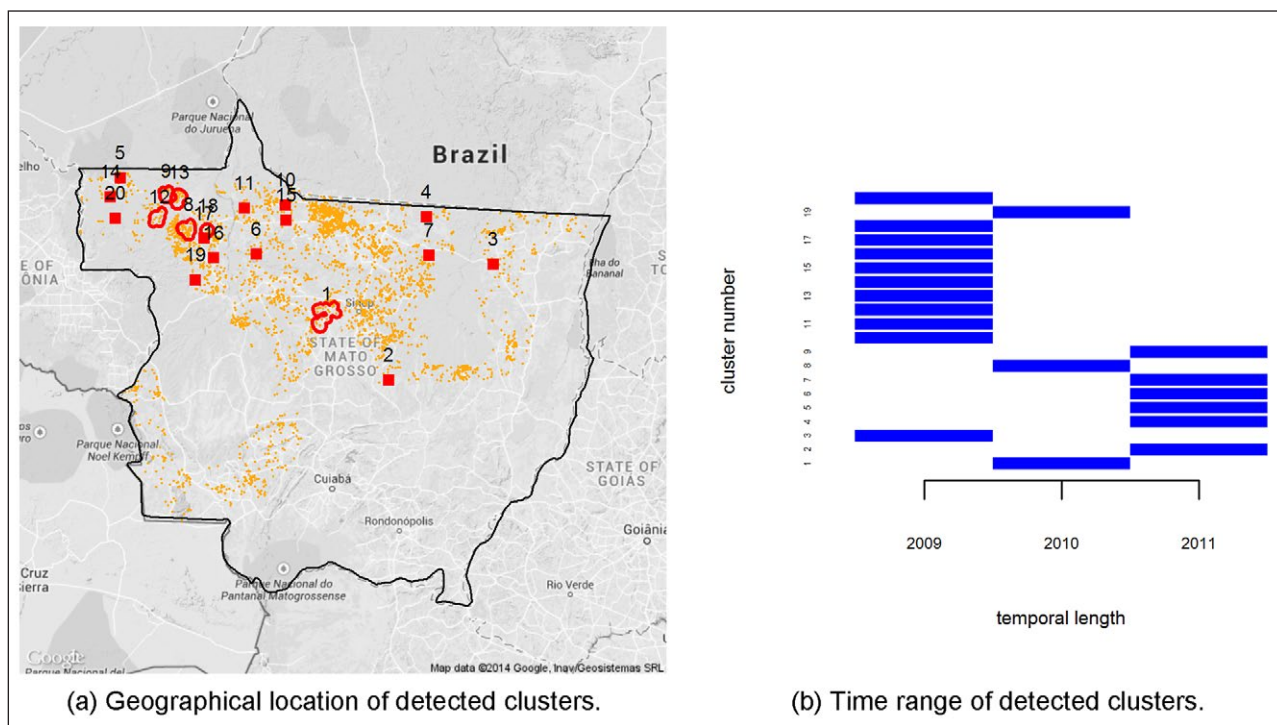
Figure 6: Spatial clustering analysis for the state of Mato Grosso (MT). (a) Detected clusters in the state of Mato Grosso are quite small. Squares represent clusters which include only one property. **(b)** Detected clusters: 2, 4, 5, 6, 7 and 8 were found to be statistically significant in 2011. Clusters 1, 8 and 19 were statistically significant in 2010, and the remaining detected clusters were found to be statistically significant in 2009. DOI: https://doi.org/10.1525/elementa.260.f6

(a) Geographical location of detected clusters.

(b) Time range of detected clusters.

Figure 7: Spatial clustering analysis for the state of Pará (PA). (a) Detected cluster in the state of Pará are larger as compared to the detect clusters in the state of Mato Grosso. DOI: https://doi.org/10.1525/elementa.260.f7

5. Discussion and conclusion

We applied statistical epidemiological models to evaluate the deforestation rates in the states of Mato Grosso and Pará, in Brazil. Standard statistical models for count data which are the Binomial, Poisson and Negative Binomial were tested and did not fit properly the data. The data shows that most properties had zero deforestation in a given year. A zero-inflated model, was also evaluated

and did not achieve a proper fit either. Thus, statistical simulation tools derived from epidemiology were developed in order to evaluate whether the governmental policy named CAR achieved its highest goal, to reduce the deforestation rates.

A statistical comparison between properties which did adopt CAR and those which did not adopted CAR in the studied period showed that the properties inside the Rural Environmental Registry have reduced their deforestation rates in some property classes and time periods, but this effect has not been systematic across time and space. This indicate that the effectiveness of CAR in reducing deforestation was only partial. For small properties, CAR seemed to have a stronger effect during the initial years of implementation but this result faded during time. For medium and high properties, alternating higher results between CAR and without CAR properties suggest that other factors than CAR may be influencing the deforestation dynamics.

Space-time cluster analysis were applied to the data in order to detect areas and time periods in which the deforestation rates were higher than the expected rate under the null hypothesis of space-time randomness. Larger clusters were found in state of Pará and smaller clusters were found in the state of Mato Grosso. The Poisson model was used in the cluster analysis but since previous statistical analysis had revealed that the Poisson model did not fit the data properly, the results represent exploratory analysis. Future works aim at developing a cluster analysis using a proper statistical model which accounts for over dispersion in the data.

As opposed to purely spatial analysis, space-time cluster analysis may indicate live clusters, i.e., clusters which comprise the last year of data, indicating that, in these clusters, deforestation may still happen in following years. Furthermore, clusters which do not comprise the last year of data indicate properties in which deforestation had decreased significantly. Results indicate smaller clusters of deforestation in the state of Mato Grosso and larger clusters in the state of Pará. Results also show live clusters in both states indicating a continuous deforestation process in some groups of properties.

This study has both policy and methodological implications. On the policy front, it indicates that CAR has not been able to reduce deforestation across the entire period and property sizes, reducing its effectiveness in small properties over time. This highlights the need of not only incentivizing farmers to join the registry but also actively use it to tackle illegal deforestation, and inform the population about the increased monitoring capabilities of the government in order to avoid deforestation. From a methodological point of view, the proposed statistical models contain some advantages over the econometric and simulation models that are currently widely applied to evaluate environmental policies. In contrast to econometric models that require detailed economic and social data (Angelsen, 1999; Pfaff et al., 2008), the proposed statistical models can be applied to situations in which such data is not available. The proposed methods also use the entire dataset without the

need to calculate the difference in differences between matched subsets of samples. Similarly, the proposed statistical method provides some advantages in relation to simulation methods since it gives results that rely less on the modeler's design choices or on the assumptions of specific statistical distributions. For these reasons, we believe that the present statistical models may find a wide application of similar policy evaluation problems, especially in data-poor contexts in developing countries.

Data Accessibility Statement
The data can be requested by contacting macosta@ufmg.br.

Supplemental File
The supplemental file for this article can be found as follows:

- **DataRCode.** Data S1. Rar. DOI: https://doi.org/10.1525/elementa.260.s1

Competing interests
The authors have no competing interests to declare.

Author contributions
- Contributed to conception and design: MAC, RR, MCCS, AAA
- Contributed to acquisition of data: MAC, MCC, AAA
- Contributed to analysis and interpretation of data: MAC, RR, MCCS, AAA
- Drafted and/or revised the article: JC
- Approved the submitted version for publication: MCC, RR, JC

References
Angelsen, A 1999 Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights. *J Dev Econ* **58**(1): 185–218. DOI: https://doi.org/10.1016/S0304-3878(98)00108-4

Arima, EY, Barreto, P, Araújo, E and **Soares-Filho, B** 2014 Public policies can reduce tropical deforestation: Lessons and challenges from Brazil. *Land Use Policy* **41**(0): 465–473. DOI: https://doi.org/10.1016/j.landusepol.2014.06.026

Azevedo, AA, Rajão, R, Costa, MA, Stabile, MCC, Macedo, MN, dos Reis, TNP, Alencar, A, Soares-Filho, BS and **Pacheco, R** 2017 Limits of Brazil's Forest Code as a means to end illegal deforestation. *Proc Natl Acad Sci* **114**(29): 7653–7658. DOI: https://doi.org/10.1073/pnas.1604768114

Campbell, DT and **Stanley, JC** 1966 *Experimental and Quasi Experimental Designs for Research*. Reprinted from Handbook of Research on Teaching, 1963. Boston, NY, USA: Houghton Mifflin Company.

Costa, MA and **Kulldorff, M** 2014 Maximum linkage space-time permutation scan statistics for disease outbreak detection. *Int J Health Geogr* **13**(1): 20. DOI: https://doi.org/10.1186/1476-072X-13-20

Coulston, JW and **Riitters, KH** 2003 Geographic Analysis of Forest Health Indicators Using Spatial Scan Statistics. *Environ Manage* **31**(6): 764–773. DOI: https://doi.org/10.1007/s00267-002-0023-9

Crabbé, A and **Leroy, P** 2008 *The Handbook of Environmental Policy Evaluation.* London, UK: Earthscan.

Dekking, FM 2005 *A Modern Introduction to Probability and Statistics: Understanding Why and How.* Springer Science & Business Media. DOI: https://doi.org/10.1007/1-84628-168-7

Dwass, M 1957 Modified Randomization Tests for Nonparametric Hypotheses. *Ann Math Stat* **28**(1): 181–187. DOI: https://doi.org/10.1214/aoms/1177707045

Efron, B 1979 Bootstrap Methods: Another Look at the Jackknife. *Ann Stat* **7**(1): 1–26. DOI: https://doi.org/10.1214/aos/1176344552

Efron, B and **Tibshirani, RJ** 1994 *An Introduction to the Bootstrap.* CRC press.

Fei, S n.d. Applying Hotspot Detection Methods in Forestry: A Case Study of Chestnut Oak Regeneration. *Int J For Res* **2010**: 8. DOI: https://doi.org/10.1155/2010/815292

Ferraro, PJ 2009 Counterfactual thinking and impact evaluation in environmental policy. *New Dir Eval* **2009**(122): 75–84. DOI: https://doi.org/10.1002/ev.297

Ferraro, PJ and **Miranda, JJ** 2014 The performance of non-experimental designs in the evaluation of environmental programs: A design-replication study using a large-scale randomized experiment as a benchmark. *J Econ Behav Organ* **107**(Part A(0)): 344–365. DOI: https://doi.org/10.1016/j.jebo.2014.03.008

Gibbs, HK, Rausch, L, Munger, J, Schelly, I, Morton, DC, Noojipady, P, Soares-Filho, B, Barreto, P, Micol, L and **Walker, NF** 2015 Brazil's Soy Moratorium. *Science* **347**(6220): 377–378. DOI: https://doi.org/10.1126/science.aaa0181

Glasserman, P 2003 *Monte Carlo Methods in Financial Engineering.* Springer Science & Business Media **53**. DOI: https://doi.org/10.1007/978-0-387-21617-1

Godar, J, Gardner, TA, Tizado, EJ and **Pacheco, P** 2014 Actor-specific contributions to the deforestation slowdown in the Brazilian Amazon. *Proc Natl Acad Sci* **111**(43): 15591–15596. DOI: https://doi.org/10.1073/pnas.1322825111

Greenstone, M and **Gayer, T** 2009 Quasi-experimental and experimental approaches to environmental economics. *J Environ Econ Manag* **57**(1): 21–44. DOI: https://doi.org/10.1016/j.jeem.2008.02.004

Jewell, NP 2003 *Statistics for Epidemiology.* CRC Press.

Kulldorff, M 1997 A spatial scan statistic. *Commun Stat – Theory Methods* **26**(6): 1481–1496. DOI: https://doi.org/10.1080/03610929708831995

Kulldorff, M, Athas, WF, Feurer, EJ, Miller, BA and **Key, CR** 1998 Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am J Public Health* **88**(9): 1377–1380. DOI: https://doi.org/10.2105/AJPH.88.9.1377

Lawson, AB 2013 *Statistical Methods in Spatial Epidemiology.* John Wiley & Sons.

Nelder, JA and **Baker, RJ** 1972 *Generalized Linear Models.* Wiley Online Library. Available at: http://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess0866.pub2/full Accessed 2017 Oct 4.

Nepstad, D, Schwartzman, S, Bamberger, B, Santilli, M, Ray, D, Schlesinger, P, Lefebvre, P, Alencar, A, Prinz, E, Fiske, G, et al. 2006 Inhibition of Amazon Deforestation and Fire by Parks and Indigenous Lands. *Conserv Biol* **20**(1): 65–73. DOI: https://doi.org/10.1111/j.1523-1739.2006.00351.x

Pfaff, A, Robalino, JA, Sanchez-Azofeifa, GA 2008 Payments for environmental services: empirical analysis for Costa Rica. *Terry Sanford Inst Public Policy Duke Univ Durh NC USA*, in press.

Ribeiro, SHR and **Costa, MA** 2012 Optimal selection of the spatial scan parameters for cluster detection: A simulation study. *GEOMED 2011 Spec Issue* **3**(2): 107–120. DOI: https://doi.org/10.1016/j.sste.2012.04.004

Rosa, IMD, Purves, D, Souza, C, Jr. and **Ewers, RM** 2013 Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. *PLoS ONE* **8**(10): 1–14. DOI: https://doi.org/10.1371/journal.pone.0077231

Selvin, S 2004 *Statistical Analysis of Epidemiologic Data* **3**. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195172805.001.0001

Shadish, WR, Cook, TD and **Campbell, DT** 2001 *Experimental and Quas-Experimental Designs for Generalized Causal Inference.* 2nd Edition. Boston, USA: Cengage Learning.

Smith, EP 2002 BACI Design. In: El-Shaarawi, AH and Piegorsch, WW (eds.), *Encyclopedia of Environmetrics,* 141–148. Chichester: John Wiley & Sons LTD.

Soares-Filho, B, Moutinho, P, Nepstad, D, Anderson, A, Rodrigues, H, Garcia, R, Dietzsch, L, Merry, F, Bowman, M, Hissa, L, et al. 2010 Role of Brazilian Amazon protected areas in climate change mitigation. *Proc Natl Acad Sci* **107**(24): 10821–10826. DOI: https://doi.org/10.1073/pnas.0913048107

Soares-Filho, B, Rajão, R, Macedo, M, Carneiro, A, Costa, W, Coe, M, Rodrigues, H and **Alencar, A** 2014 Cracking Brazil's Forest Code. *Science* **344**(6182): 363–364. DOI: https://doi.org/10.1126/science.1246663

Soares-Filho, BS, Nepstad, DC, Curran, LM, Cerqueira, GC, Garcia, RA, Ramos, CA, Voll, E, McDonald, A, Lefebvre, P and **Schlesinger, P** 2006 Modelling conservation in the Amazon basin. *Nature* **440**(7083): 520–523. DOI: https://doi.org/10.1038/nature04389

Stewart-Oaten, A, Murdoch, WW and **Parker, KR** 1986 Environmental Impact Assessment: "Pseudoreplication" in Time? *Ecology* **67**(4): 929–940. DOI: https://doi.org/10.2307/1939815

Tonini, M, Tuia, D and **Ratle, F** 2009 Detection of clusters using space–time scan statistics. *Int J Wildland Fire* **18**(7): 830–836. DOI: https://doi.org/10.1071/WF07167

Tuia, D, Ratle, F, Lasaponara, R, Telesca, L and **Kanevski, M** n.d. Scan statistics analysis of forest fire clusters. *Commun Nonlinear Sci Numer Simul* **13**(8): 1689–1694. DOI: https://doi.org/10.1016/j.cnsns.2007.03.004

Vega Orozco, C, Tonini, M, Conedera, M and **Kanveski, M** 2012 Cluster recognition in spatial-temporal sequences: the case of forest fires. *GeoInformatica* **16**(4): 653–673. DOI: https://doi.org/10.1007/s10707-012-0161-z

Wedderburn, RW 1974 Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**(3): 439–447.

Wiens, JA and **Parker, KR** 1995 Analyzing the Effects of Accidental Environmental Impacts: Approaches and Assumptions. *Ecol Appl* **5**(4): 1069–1083. DOI: https://doi.org/10.2307/2269355