

BRAZILIAN AMAZONIA DEFORESTATION DETECTION USING SPATIO-TEMPORAL SCAN STATISTICS

C. A. O. Vieira^{a,*}, N. T. Santos^b, A. P. S. Carneiro^b, A. A. S. Balieiro^c

^a Dept. Geoscience, Federal University of Santa Catarina, Trindade, Florianópolis – SC, Brazil
carlos.vieira@ufsc.br

^b Federal University of Viçosa, Viçosa - MG, Brazil
[\(nsantos, policarpo\)@ufv.br](mailto:(nsantos, policarpo)@ufv.br)

^c Fundação de Vigilância em Saúde do Estado Amazonas - FVS/AM
Av. André Araújo nº 701 - Aleixo, CEP. 69.060-001 - Manaus - Amazonas, Brasil
alcirley@gmail.com

KEY WORDS: Data Mining, Forestry, Change Detection, Remote Sensing, Spatial, Statistics

ABSTRACT:

The spatio-temporal models, developed for analyses of diseases, can also be used for others fields of study, including concerns about forest and deforestation. The aim of this paper is to quantitatively check priority areas in order to combat deforestation on the Amazon forest, using the space-time scan statistic. The study area location is at the south of the Amazonas State and cover around 297.183 kilometre squares, including the municipality of Boca do Acre, Labrea, Canutama, Humaita, Manicore, Novo Aripuana e Apuí County on the north region of Brazil. This area has showed a significant change for land cover, which has increased the number of deforestation's alerts. Therefore this situation becomes a concern and gets more investigation, trying to stop factors that increase the number of cases in the area. The methodology includes the location and year that deforestation's alert occurred. These deforestation's alerts are mapped by the DETER (Detection System of Deforestation in Real Time in Amazonia), which is carry out by the Brazilian Space Agency (INPE). The software SatScanTM v7.0 was used in order to define space-time permutation scan statistic for detection of deforestation cases. The outcome of this experiment shows an efficient model to detect space-time clusters of deforestation's alerts. The model was efficient to detect the location, the size, the order and characteristics about activities at the end of the experiments. Two clusters were considered actives and kept actives up to the end of the study. These clusters are located in Canutama and Lábrea County. This quantitative spatial modelling of deforestation warnings allowed: firstly, identifying actives clustering of deforestation, in which the environment government official are able to concentrate their actions; secondly, identifying historic clustering of deforestation, in which the environment government official are able to monitoring in order to avoid them to became actives again; and finally, verify that distances between the deforestation warning and the roads explain part of the significant clustering.

1. INTRODUCTION

There is a high biodiversity at the Amazonian forest. Environmental problems on the Amazonian Forest are usually related to deforestation and fire. The Amazonian State is less affected by deforestation, than the other ones; however, there are some deforestation problems on the south boundary of the State.

The Brazilian Space Agency (INPE) conducts project that monitor Amazonian deforestation, such as: DETER (Detection System of Deforestation in Real Time in Amazonia). However, even identifying deforestation areas using remotely sensed data in the Amazon Forest, these projects are not able to verify quantitatively the magnitude and establish the priority of monitoring theses deforestation clusters. Thus, there is a lack of spatial analysis in order to model quantitatively these deforestation clusters.

The spatio-temporal statistical models, developed for diseases analyses (Kulldorff et al., 2005), can also be used for others subjects, including concerns about forest management and

deforestation. The aim of this paper is to quantitatively check priority areas in order to combat deforestation on the Amazon forest, using the space-time scan statistic proposed by Kulldorff (1995). Thus, this paper uses a methodology for detecting space-time clusters of deforestation cases that were mapped through the investigation of deforestation in Amazonas State.

2. STUDY AREA AND MATERIALS

2.1 Study Area

The study area location is at the south of the Amazonas State, including the municipalities of Boca do Acre, Labrea, Canutama, Humaita, Manicore, Novo Aripuana e Apuí County (Figure 1). This area has showed a significant change for land cover, which has increased the number of deforestation's alerts. The study area covers around 297.183 km².

The region is above the sea levels, with altitude varying from 40 m to 150 m. The vegetation is characterized as a deep

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

tropical forestry. The climate is humid equatorial with temperature on average from 17⁰C to 35.5⁰C. The region presents two very well defined seasons: winter (rain season) and summer (dry season). Two rivers split up the regions: Purus River and Madeira River, and their effluents (IBGE, 2008).

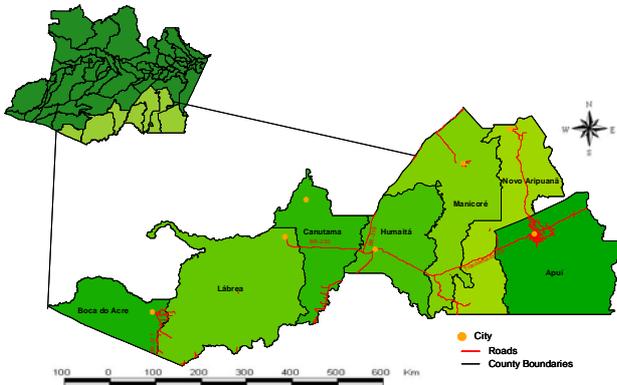


Figure 1. Location of the Amazon State and the cities of: Boca do, Lábrea, Canutama, Humaitá, Manicoré, Novo Aripuanã and Apuí.

2.2 Dataset

The methodology includes the location and year that deforestation's alert occurred. These deforestation's alerts are mapped by the DETER (Detection System of Deforestation in Real Time in Amazonia), which is carry out by the Brazilian Space Agency (INPE). The software SatScanTM v7.0 was used in order to define space-time permutation scan statistic for detection of deforestation cases. Other systems used in the pre-processing data were Excel, ARCVIEW 3.2 and R 2.7.1. It was also used maps in shape format with county boundaries (IBGE, 2008), and other shapes files with official roads (SISCOM, 2008). Tables with warnings of deforestation from 2004 to 2007 (INPE-DETER, 2008) derived from remotely sensed data (MODIS). Finally, statistic test were carried out in order to check the hypothesis with 5% probability.

3. METHODS

Procedures for dataset preparations and details of the spatial-temporal models are described on this section.

3.1 Dataset preparations

Tables with deforestations warning from DETER project had coordinates points that identified deforestation centres (centroid). However, these databases were not complete for the whole set of districts, because there were clouds, which blocked the remotely sensed monitoring. The software SATSCAN requires that the temporal information be aggregated for a period of time, and it was chose to aggregate the database per year, as suggested by INPE (2008). It was input into an Excel table of the deforestation warnings. From these tables were extracted the variables and co-variables in order to generate the case and coordinate files, both following *.TXT format. These TXT files were processed by the software SATSCAN and the results were spatialized using the ARCVIEW and the software R.

In order to verify the existence of clustering, the spatial-temporal permutation model proposed by Kulldorff et al. (2005) was utilized. Once that the database was composed only by "cases", or the number of points with deforestation (variable) on the study area. The population at risk is the forest. As there is evidence that distances from roads contributed to increase deforestation areas (Brandão Jr et al., 2007), it was chosen the distance from the roads in this model as categorical co-variable. The distances to the nearest road were classified arbitrarily into 6 categories. Thus, every case on the database received an attribute from 1 to 6, indicating the distance from roads. For instance, cases were attributed value 1 if the warning happens between 0 – 10 km from roads; 2 if between 10 – 20 km; and so on; and finally, 6 if higher than 50 km from the roads. An accessibility map (a map of the distance from the nearest roads) can be shown in Figure 2, for the later comparison of retrospective analysis with and without the distance from roads.

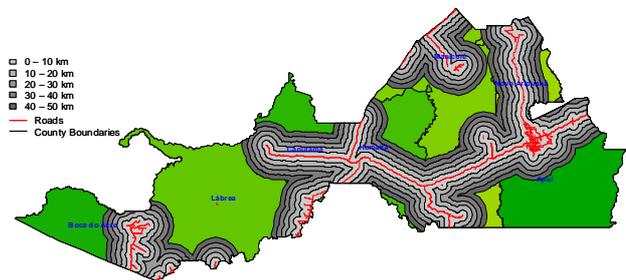


Figure 2. Accessibility map (a map of the distance from the nearest roads).

The radius of search was set up to 20 km following Riitters & Coulston (2005) suggestion, in order to deforestation search, on the west of USA. It was also adopted a backward analysis, in order to detect not only the active clusters, the ones that last up the end of the study, but also the historical clusters, the ones which disappeared before finish the study period.

3.2 The Spatial-Temporal Model

The spatio-temporal model assumed that the deforestation cases are distributed according to Hipergeometric distribution. This model could be formalized as following: be c_{zt} the observed number of cases at the geographic area z , during determined time t . The total number of deforestation warnings (C), could be defined by:

$$C = \sum_z \sum_t c_{zt} \quad (1)$$

In addition, the expected number of deforestation warning (μ_A) inside determined cylinder A , is defined by:

$$\mu_A = \frac{1}{C} \left(\sum_{t \in \Omega} \sum_{z \in A} c_{zt} \right) \left(\sum_{z \in \Omega} \sum_{t \in A} c_{zt} \right) \quad (2)$$

The clusters had a cylinder shape. Each one of this clusters were obtained through variations on spatial radius and time interval of search (see Figure 3). Note that first part of the Equation (2) – in green, corresponds to the deforestation cases inside the area A of the cylinder in Figure 3; and, the second

part of Equation (2) - in blue, corresponds to the deforestation cases that occurred inside the time t .

A Generalized Likelihood Ratio of Poisson (RVG), in Equation (3), was obtained for every candidate cluster.

$$RVG = \left(\frac{c_A}{\mu_A} \right)^{c_A} \left(\frac{C - c_A}{C - \mu_A} \right)^{(C - c_A)} \quad (3)$$

where, C is the total number of deforestation warnings from 2004 to 2007 for the cities of: Apuí, Novo Aripuanã, Manicoré, Humaitá, Canutama, Lábrea and Boca do Acre; c_A is the number of deforestation warning inside determined cylinder A ; μ_A is the expected number of deforestation warning inside determined cylinder A .

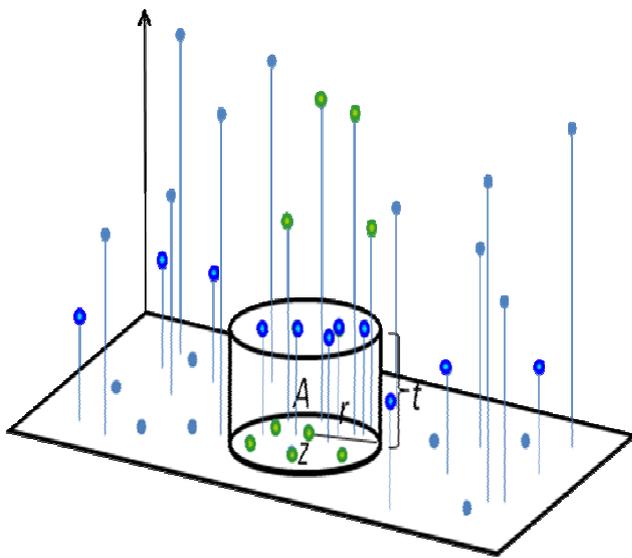


Figure 3. Cluster illustration of the spatial-temporal model

Considering the inclusion of the co-variable distance from roads, the RGV used is slightly changed to:

$$RGV = \prod_{i=1}^6 \left(\frac{c_{A(i)}}{\mu_{A(i)}} \right)^{c_{A(i)}} \left(\frac{C_i - c_{A(i)}}{C_i - \mu_{A(i)}} \right)^{C_i - c_{A(i)}} \quad (4)$$

where, i is one of the 6 categories of the co-variable attribute distance from roads. If determined co-variable explains a cluster or part of it, then this cluster would disappear or reduce, during this analysis. The likely cluster, said primary, is the one that presented the higher RGV value. Suppose that the order of the RGV of this primary cluster be equal to R , in order to test the significance of this cluster, one could set up the hypothesis:

H_0 : There is no cluster spatio-temporal of the number of deforestation warnings; against,

H_1 : There exist cluster spatio-temporal of the number of deforestation warning on the study area.

In order to test these hypothesis with relation to the primary cluster, the *Monte Carlo* simulation was performed, which consisted in simulate 999 permutations, at random, of the number of case in relation to the area and time under analysis, keeping unchanged the spatial and temporal marginal. Figure 4 illustrates Monte Carlo Simulations procedures.

Monte-Carlo Simulations

City	2004	2005	2006	2007	Total
Apuí	1	0	0	0	1
Apuí	0	0	1	0	1
Bocado Acre	0	1	0	0	1
:	:	:	:	:	:
Humaitá	1	0	0	0	1
Lábrea	0	0	1	0	1
Manicoré	0	0	0	1	1
Novo Aripuanã	0	1	0	0	1
Total	396	302	345	267	1310

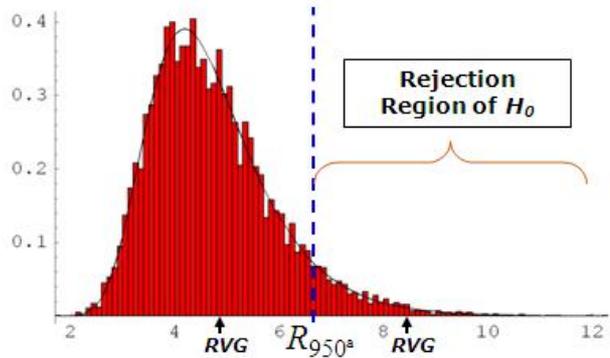


Figure 4. Monte Carlo procedure Illustration

For every one of these permutations, it was obtained the RGV values for all candidates to cluster. The maximum RGV values were used in order to obtain the ordinated distribution. The significance of the primary cluster was observed identifying in which position its order R is located in this RGV distribution. If R be between the $100(1 - \alpha)\%$ higher posts, then reject H_0 , on the significance level α . Whether R was above the 950^a position, then it is possible to conclude that the number of deforestation warnings on that cluster do not happen at random. The detected clusters, in the database, which presented values between the cluster said primary and above the 950^a position, were identified as secondary clusters. The p -value can be computed using the Equation (5):

$$p\text{-value} = 1 - R / (S + 1) \quad (5)$$

4. RESULTS AND DISCUSSION

In the Table 1, it is shown the significant cluster using just the variable numbers of deforestation warning. It was found 8 significant clusters (p -value < 0.05) from the total of 28 detected clusters. The primary cluster was localized on the county of Apuí, with radius of 19.50 km, and the deforestation centre (centroid) with geographic coordinates of 7.27S and 59.86W (Table 1).

Clusters	Period	Location/ Radius	N.º observed warning	N.º expected warning	RGV	<i>p-value</i>
<i>P</i>	2004	(7,27S; 59,86 W) / 19,50 km <i>Apuí</i>	51	19,95	17,19	0,001
<i>S1</i>	2007	(8,77S; 66,80 W) / 17,90 km <i>Lábrea</i>	52	22,22	14,79	0,001
<i>S2</i>	2006	(9,71S; 66,69 W) / 19,97 km <i>Lábrea</i>	32	12,11	11,35	0,001
<i>S3</i>	2006	(8,92S; 64,56 W) / 15,59 km <i>Lábrea, Canutama</i>	28	10,53	10,02	0,001
<i>S4</i>	2005	(9,50S; 67,06 W) / 17,36 km <i>Lábrea</i>	17	5,07	8,69	0,001
<i>S5</i>	2004	(7,86S; 61,55 W) / 15,18 km <i>Manicoré</i>	29	13,00	7,36	0,008
<i>S6</i>	2004	(6,90S; 60,09 W) / 14,91 km <i>Apuí, Novo Aripuanã</i>	13	3,93	6,51	0,038
<i>S7</i>	2007	(8,59S; 64,36 W) / 18,90 km <i>Lábrea, Canutama</i>	18	6,73	6,49	0,040

Table 1. Retrospective analysis without use of co-variable distance from roads.

In the area of this cluster, for the year of 2004, the numbers of deforestation warnings were equal to 51. Under the hypothesis of nullity, i.e., there is no cluster of the number of warning in this scanning region. It would be expected just 19.95 deforestation warning. The *p-value* for this cluster was obtained using the 999 *Monte Carlo* simulations, was equal to 0.001. This *p-value* indicates that the probability of obtain the number of warning clustered at random is smaller than 0.01%. As this value is smaller the 5%, the hypothesis of nullity is rejected. Therefore, the primary cluster was statistically significant. The significance of the secondary clusters was identified following the same procedure in order to identify the primary cluster.

On the region of the primary cluster is localized the road BR-320, one of the most important road to access the Amazon State (Figure 5). The total cluster number detected was 30 clusters. This primary cluster (*P*) is historical, i.e., it occurred in 2004, and do not last active up to 2007. Two clusters were considered actives, i.e., it continues in activity up to the end of this study (2007). The secondary 01 (*S1*) cover a region with radius of 17.90 km, with geographic coordinates of 8.77S and 66.80W, located on the county of Lábrea. The total number of observed cases was 52 against 22.22 under a nullity hypothesis. The cluster secondary 07 (*S7*) is located on the counties of Canutama and Lábrea. The major part of this cluster is located on the county of Canutama. Its area cover a radius of 18,90 km with geographic coordinates of 8.59S and 64.36W. The number of observed deforestation warning was 18, and it was expected 6.73 warning under the nullity hypothesis. The clusters remain are historical.

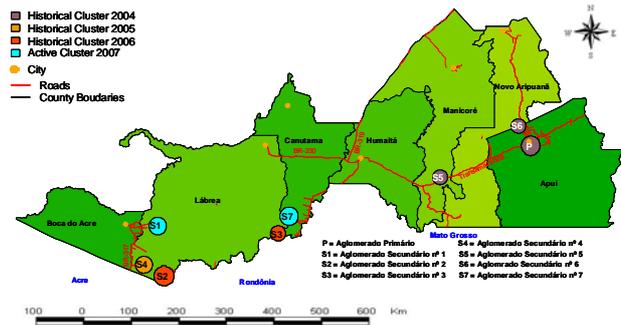


Figure 5. Retrospective analysis without use of co-variable distance from roads

In the Table 2 are shown the significant clusters using the variable number of warnings and the co-variable distance from the roads. Considering that there is reports stating that deforestation case associated to distance to roads (Brandão Jr. et al., 2007; Silva, 2006), the results from the roads explain four clusters, or part of them, when including this co-variable. There was a change on the scenario previously observed, without consider the distance from roads (Table 1).

The first fact observed was the primary cluster location is in Lábrea (and not in Apuí) and radius equal to 7.12 km (different of 19.50 km). Another fact important to mention is a reduced number of significant clusters (Figure 6), which confirm that distances between the deforestation warning and roads explain part of the significant clustering.

5. CONCLUSIONS

The outcome of this experiment shows an efficient model to detect space-time clusters of deforestation's alerts. The model was efficient to detect the location, the size, the order and characteristics about activities at the end of the experiments. Two clusters were considered actives and kept actives up to the end of the study. These clusters are located in Canutama and Lábrea County. This quantitative spatial modelling of deforestation warnings allowed: 1) Identify actives clustering of deforestation, in which the environment government official are able to concentrate their actions; 2) Identify historic clustering of deforestation, in which the environment government official are able to monitoring in order to avoid them to became actives again; and finally, 3) Verify that distances between the deforestation warning and the roads explain part of the significant clustering.

Cluster s	Period	Location/ Radius	N.º observed warning	N.º expected warning	RGV	P- value
P	2007	(8,74S;66,89W) / 7,12 km <i>Lábrea</i>	25	6,50	15,30	0,001
S1	2006	(9,71S; 6,69W) / 19,97 km <i>Lábrea</i>	32	11,12	13,12	0,001
S2	2004	(7,27S; 59,86W) / 19,50 km <i>Apuí</i>	51	23,46	12,36	0,001
S3	2007	(8,68S; 64,12W) / 14,11 km <i>Canutama</i>	22	8,54	7,43	0,005
S4	2005	(9,50S; 67,06 W) / 17,36 km <i>Lábrea</i>	17	5,84	7,05	0,011
S5	2005	(7,83S; 61,35W) / 17,83 km <i>Manicoré, Novo Aripuanã</i>	24	10,47	6,45	0,023

Table 2. Retrospective analysis considering the distance to roads (co-variable)

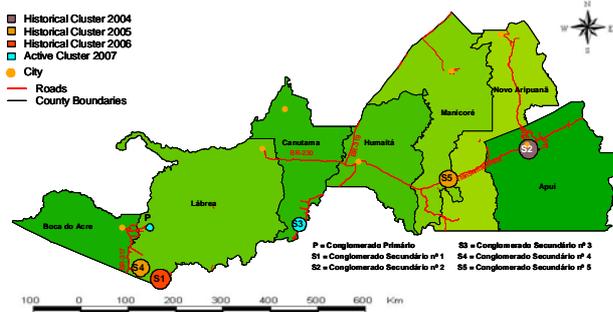


Figure 6. Retrospective analyses considering the distance to roads (co-variable)

ACKNOWLEDGEMENTS

Authors acknowledge Fundação de Amparo a Pesquisa no Estado do Amazonas (FAPEAM) for given scholarship. We also owe thanks to INPE through Dr. Yosio Shimabukuro and Sr. Egídio Arai, for the MODIS Imagery. We would like to thank to Professor Marcelo Azevedo Costa and Martin Kulldorff for their support.

REFERENCES

Aguiar, A. P. D., Câmara, G., Escada, M. I., Amaral, S., Carneiro, T., Monteiro, A. M. V., Araújo, R., Vieira, I., Becker, B., 2008. Amazon deforestation models: challenging the only-roads approach. http://www.dpi.inpe.br/gilberto/lucc/amazon_deforestation_not_only_roads.pdf (10th May 2008).

Arima, E., Brito, M., Barreto, P., 2005. Pecuária e desafios para a conservação ambiental na Amazônia. *O Estado da Amazônia*, nº 5. Belém: Imazon. 4p.

Brandão Jr, A. O., Souza Junior, C. M., Ribeiro, J. G. F., Sales, M. H. R., 2007. Desmatamento e estradas não oficiais da

Amazônia. In: *Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, 04, Florianópolis. Anais... São José dos Campos: INPE, 2007. Artigos, p. 2357-2364. CD-ROM, On-line. ISBN: 978-85-17-00031-7. <http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.15.19.47/doc/2357-2364.pdf> (20th Feb 2008).

Fearnside, P. M., 2002. Fogo e emissão de gases de efeito estufa dos ecossistemas florestais da Amazônia brasileira. *Estudos Avançados*, v. 16(44), p. 99-123.

IBGE. 2008. Download – Geociências – Mapas – Malha digital. http://www.ibge.gov.br/servidor_arquivos_geo/ (10th Mar 2008).

INPE-DETER. 2008. <http://www.obt.inpe.br/deter/> (10th Mar 2008).

Kulldorff, M. e Nagarwalla, N., 1995. Spatial disease cluster: detection and inference. *Statistics in Medicine*. v. 14, p.799-810.

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. M., Mostashari, F., 2005. A space-time permutation Scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, 2:216-224.

Kulldorff, M. and Information Management Services, Inc. 2006. SatScanTM v7.0: Software for the spatial and spacetime Scan statistics. <http://www.satscan.org/> (10th Mar 2008).

Mesquita Jr, H. N., Silva, M. C., Watanabe, N. Y., Esteves, R. L., 2007. Aplicações de sensoriamento remoto para monitoramento do desmatamento da Amazônia. In: *Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, 04, Florianópolis. Anais... São José dos Campos: INPE, 2007. Artigos, p. 6835-6842. CD-ROM, On-line. ISBN: 978-85-17-00031-7. <http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.15.23.56.56/doc/6835-6842.pdf>. (20th Feb 2008).

Riitters, K. H., Coulston, J. W., 2005. Hot spots of perforated forest in the eastern United States. *Environmental Management* vol. 35(4), p. 483-492.

R Development Core Team., 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>, version 2.7.1. (10th Mar 2008).

Silva, A. N., 2006. Análise das relações entre o desflorestamento e o potencial agropecuário das terras na Amazônia. 55 p. (INPE-15177-TDI/1293) Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

SISCOM – Sistema Compartilhado de Informações Ambientais, 2008. Centro de Sensoriamento Remoto – CSR, Centro de Monitoramento Ambiental do IBAMA – CEMAM. <http://siscom.ibama.gov.br/shapes/> (20th Mar 2008).