Data mining applied to landscape pattern analysis: A study case in center-north of Rondônia state, Brazil

Mikhaela A. J. S. Pletsch[1]

[1]Departamento de Processamento de Imagens
Instituto Nacional de Pesquisas Espaciais (INPE) – São José dos Campos, SP – Brazil

mikhaela.pletsch@inpe.br

Abstract. *Tropical forests provide several benefits to human being and considering its importance, its protection measures has been included in governmental decision making. Thus, a Brazilian initiative to monitor Amazon Forest was created, PRODES project. However, it is essential to analyze the data in order to identify and avoid elements that can degrade even more this natural asset. A possibility for that it through landscape metrics. In this context, the aim of this work was to automate a previous landscape pattern analysis in center-north Rondônia state. Although the use of the software GeoDMA was satisfactory in the process, there are some important points, which should be analyzed before the application of the final decision tree to subsequent years.*

Resumo. *Florestas tropicais provêm diversos benefícios ao homem e devido à sua importância, medidas para a sua proteção tem sido incluídas nas tomadas de decisões governamentais. Assim, uma iniciativa brasileira para monitorar a Floresta Amazônica foi criada, o projeto PRODES. Entretanto, é essencial a análise dos dados de modo a identificar e evitar os elementos que degradam ainda mais esse bem natural. Uma possibilidade para isso é por meio de métricas de paisagem. Nesse contexto, o objetivo deste trabalho foi automatizar uma análise de padrões de paisagens no centro-norte do estado de Rondônia. Embora o uso do software GeoDMA tenha sido satisfatório no processo, há alguns importantes apontamentos, os quais devem ser analisados antes da aplicação da árvores de decisão final para os anos subsequentes.*

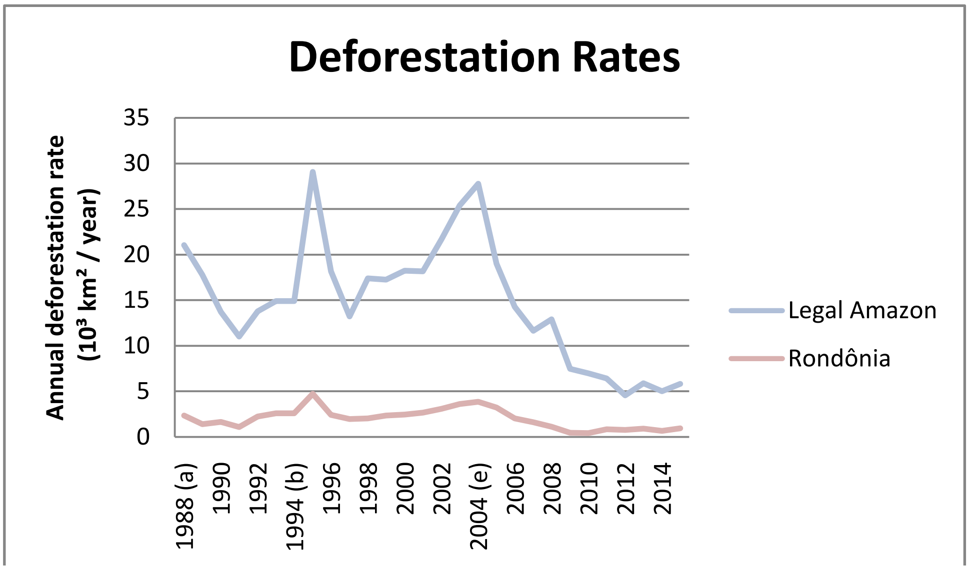
1. Introduction

Deforestation is one of the main challenges that tropical forests face and it is motivated by several complex factors (Albuquerque e Arraes, Mariano, and Simonassi 2012). According to Geist and Lambin (2001), the combination of several elements prompts this environmental degradation, as agricultural expansion, extension of infrastructure and economic tandems. Thus, through deforestation processes forests can be replaced by other land use/land cover (LULC) such as pastures, agricultural crops and settlements. Forests play an important role in providing natural resources, conservation of genetic variability and even mitigating climate change through carbon sequestration. In this context, a possibility to assure environment preservation is by the introduction of sustainable measures and establishment of an effective governance in order to avoid deforestation (Albuquerque e Arraes et al. 2012; Soares-Filho et al. 2005). However, it is hard not only characterize LULC, but also maintain the data updated (Rogan and Chen 2004). In this manner, remote sensing stands among others tools out. Its sources of data are relatively currently and efficient, which could be used to inventory and monitor these changes (Mas 1997). Taking this into account, PRODES (*Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite*), a Brazilian initiative created in 1988, aims to supervise the shallow cut deforestation in the legal Amazon through satellites (INPE, 2016a). Due to two historical documents that gave rise to the PRODES, the deforestation taxes are used by the Brazilian government since 70's. Currently operated by the Brazilian Institute of Space Research (INPE), PRODES provides annual deforestation taxes in order to support decision making.

In Brazilian Amazon, the past settlement in 50's was characterized by expansion of the agricultural frontier, with a direct result of extensive transformations. Although PRODES is an important data collection tool, a followed step, which aims its analysis, is also necessary. Thereby, landscape ecology has been employed in spatial planning, because of its spatial dimension (Aguilera, Valenzuela, and Botequilha-Leitão 2011), thus, landscape dynamics can help studies to understand deforestation process in Amazon, and factors that influences deforestation. In this context, in center-north region of Rondônia, Brazilian state, Escada (2003) locally analyzed the area and found out ten different landscape metrics. According to this author, involved actors, format, extension and age of the occupation are some of the factors that can influences LULC. To achieve this result, an empirical method partition of the space was developed, and observed spatial patterns were identified along temporal image series from the sensor TM of the Landsat satellites (TM/Landsat). Thus, the evolutionary stages of occupation process were characterized by Escada (2003) for the period 1985-2000, with temporal resolution of three years, in part of the municipalities of Cujubim, Jaru, Vale do Anari, Vale do Paraíso, Machadinho d´Oeste, Ouro Preto d´Oeste, Ariquemes, Rio Crespo, Ji-Paraná and Theobroma. In some parts, as in Machadinho D'Oeste, the initial occupation was stimulate by INCRA (National Institute for Colonization and Agrarian Reform), and along the years, different agricultural designs were developed also in its surrounds. However, the assessment process took a long time, considering that is required among others processes, field work and photo interpretation. Thus, this work developed an automation of the analysis through Knowledge Discovery in Databases (KDD), at the same time it analyzed the importance of landscape metrics of deforestation fragments and its correlation with cell sizes and deforestation patterns in Legal Amazon in center north of Rondônia. According to Fayyad and Stolorz (1997), KDD refers to the process of discovering useful knowledge from data, where data mining is one of its step. The software used was TerraView 4.2.2, and its plugin Geographical Data Mining Analyst (GeoDMA), both free distributed and developed by INPE (INPE, 2016b). Besides supporting technically future landscape ecology studies, it was also aimed through this work to develop a decision tree, which could be used in the area to support decision makers.

1.1. Deforestation rates

According to PRODES (INPE, 2016a), in the Legal Amazon as well as in Rondônia it is possible to notice two deforestation picks, 1994 and 2004 (Graph 1). As aforementioned it can be related to several complex factors (Albuquerque e Arraes et al. 2012). In the period of 1988-2015, in Legal Amazon, ca. 413500 km² were deforested, 56400 km² just in Rondônia state.



Graph 1. Deforestation in Rondônia. (a) average between 1977-1988; (b) average between 1993-1994; (e) beginning of the Plan Action for Prevention and Control of Deforestation in the Legal Amazon (PPCDAm).

1.2. Landscape ecology and deforestation patterns

Among the negative consequences that deforestation may cause on aforementioned forests benefits, this activity can also modify wildlife habitats, impacting its resources (Trani and Giles, Jr 1999). In this manner, the recognition of landscape patterns on those dynamics have become essential in many fronts. Understanding those pattern can be crucial to support decision making and to model LULC changes along the time/space and in different sceneries. In this sense, the investigation of spatial heterogeneity in various scales is the focus of landscape ecology (DiBari 2007), which properties can be quantified through a set of metrics (Aguilera et al. 2011). In other words, the characterization of spatial properties of map patterns are the objects analysis of landscape metrics (McGarigal 2013).

Husson et al. (1995) catalogued a typology of forest-non forest interfaces and fragmentation patterns that occurs in tropical forests. At first, the pattern found were: linear, insular, diffuse, massive and sub-types of them. According to it, Mertens and Lambin (1997) seek to recognize the pattern geometric, island, corridor, diffuse, fishbone and patchy in deforestation processes in southern Cameroon (Figure 1)

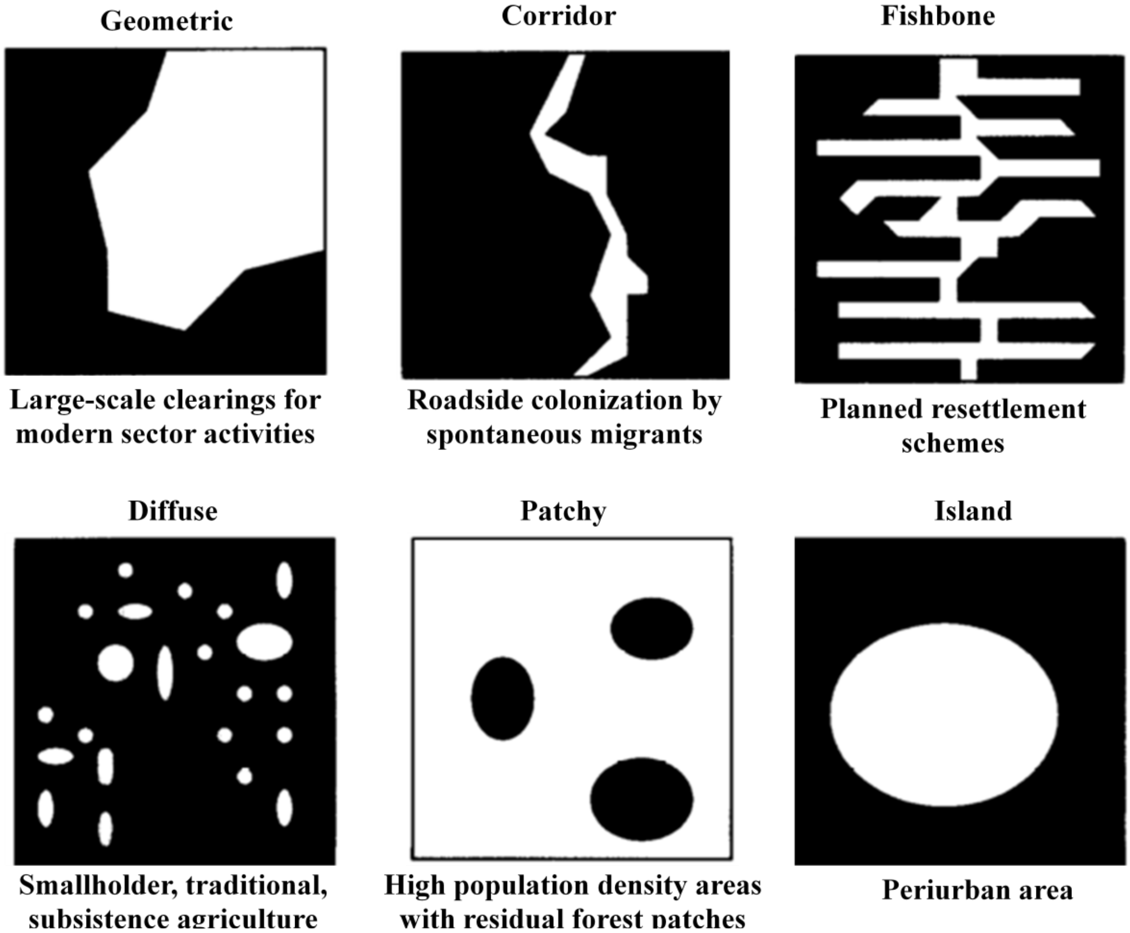


Figure 1. Typology of the forest/non-forest spatial patterns (extract from Mertens and Lambin, 1997).

In the cases from figure 1, is is possible to define a type of actor that is associated with a certain spatial configuration, and size of deforested plots. In geometric one, it is possible to associate it to large-scale clearings for modern sector activities. Along roadside it possible to identify some spontaneous migrants, corridor. Besides that, it is also possible to notice if it was planned, as fishbone, or not, as diffuse.

Based on several authors, Escada (2003) identified ten kinds of properties in center-north of Rondônia, Brazil. The area analyzed by Escada (2003) comprehends Machadinho D'Oeste, Cujubim, Jaru, Vale do Paraíso, Theobroma e Oneide and it was identified distinct rhythm and intensity LULC changes. The analysis parameters used were extension of properties, official occupation planning, occupation age, spatial configuration and plot of land extension. In its methodology, besides INCRAs occupation maps, other information were used, such as field work and farming census data. Furthermore, some questionnaires were applied to 40 land farm ownerships, looking to identify the main deforestation processes.

2. Material and Methods

To analyze the efficiency of the process automation and determine the importance of landscape metrics, firstly deforestation images from 1985-2000, with temporal resolution of three years were used. The free and qualified geoinformatic software used was TerraView 4.2.2, and its plugin GeoDMA, both distributed and developed by INPE (INPE, 2016b). This plugin, which methodology was proposed by (Silva et al. 2005) to identify deforestation patterns in Amazon, is coded in C++, and provides a interface to the interpreter with geographic information data stored in databases (Körting, Garcia Fonseca, and Câmara 2013). GeoDMA can deal with distinct geospatial data. GeoDMA has metrics that integrate polygons, cells, and images. Through image segmentation, GeoDMA creates polygons. Thus, it is provided a list of 3 feature types: i) Segmentation-based spectral features; ii) Segmentation-based spatial features; iii) Landscape-based features (INPE, 2016c). In this study case, it were used grid of cells and landscape-based features types. However, segmentation is a challenge to researches, considering that so many desired properties are hard to find, such as uniformity and homogeneity regarding to some characteristics as gray, tone and texture (Haralick and Shapiro 1985). Therefore, it was applied three different cell sizes in order to analyze which one could describe better the landscape metrics, 15 km^2, 5 km^2, 2.5 km^2. There on, it was selected the cell size of 5 km^2, considering its efficiency to describe better some local spatial pattern in comparison with the others cell sizes. Besides that, based on Escada (2003), a typology table of the forest/non-forest spatial patterns identified in center-north of Rondônia was developed (Table 1). Although Escada (2003) identified 10 typologies types (Table 2), for this work it was reduced to 5, considering pattern's frequency on the area. For those elements that are not so frequent, it may be more appropriate to identify them manually and not through KDD. For each identified pattern, some samples were selected for the machine training process, which validation were possible through Escada (2003) (Figure 2). In validation process, some geoprocessing measure have been applied, as polygon segmentation and classification. Finally, the decision tree was generated for 5 km^2 cell size in order to be applied in the automation process.

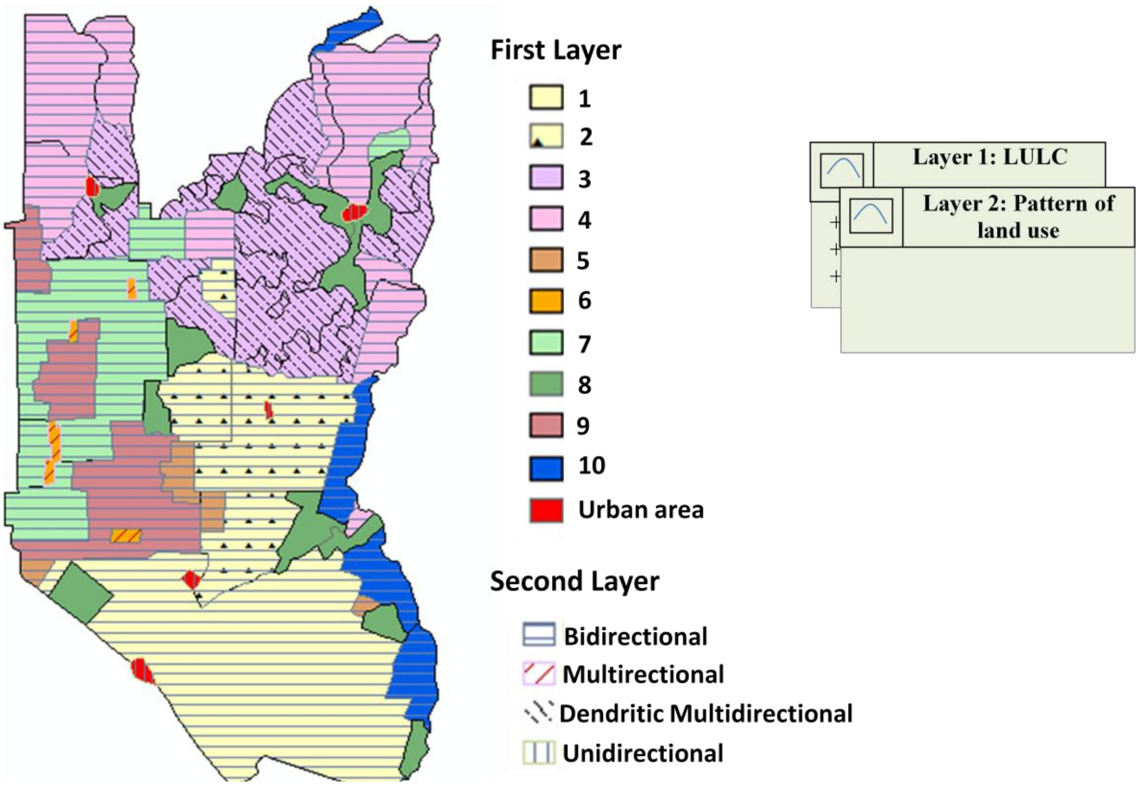
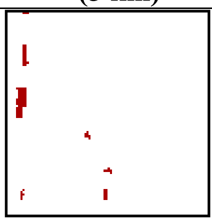
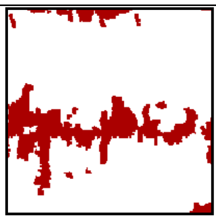


Figure 2. Original classification map - adapted from Escada (2003)

Table 1. Typology descriptions

Pattern in images	Spatial configuration (5 km)	Plot of land size	Occupation stage	Semantic
Diffuse		Small	Early	Occupation close to bank rivers and mining areas
Unidirectional / Linear		Small	Early / Intermediate	Along roads occupation. Predominantly small farmers. Related to INCRAS's settlement with different spatial distribution. Plots about 50 ha.

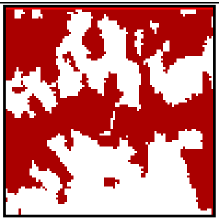
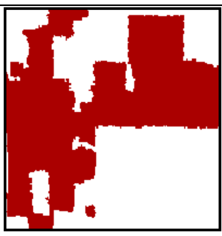
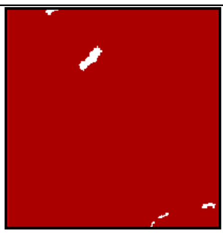
Multidirectional with dendritic pattern		Small	Intermediate	INCRA settlement areas with dendritic spatial pattern and block format reserved forest, settled between 1982 and 1984. Plots are, on average, 50 ha.
Geometric		Predominantly medium	Early / Intermediate	Predominantly medium farmers areas established over 20 years in bidding area. Plots range from 500 to 1000 ha
Consolidated		Small, medium and big	Advanced	Consolidated occupation area with small forest remnants. Areas with pioneer settlements and small farmers, established between 1970 and 1978 by INCRA. Plots of about 100 ha. Initially they may be presented as orthogonal configuration

Table 2. Original typology - Escada (2003)

Type	Description	Type	Description
1	small farmers	5	small farmers
	Settled 1971-1978		Settled 1995-1997
	about 100ha		<25ha
	Orthogonal		Orthogonal/dendritic/radial
2	small farmers	6	medium farmers
	Settled 1982-1988		bidding
	about 50ha		Small dimensions
	Orthogonal	7	medium farmers
3	small farmers		500-1000ha
	Settled 1982-1984	8	medium farmers

	about 50ha		Plots concentrations
	Dendritic	9	Big farmers
4	small farmers		>1000ha
	Settled 1990-1995	10	Big farmers
	about 50ha		Isolated areas
	Orthogonal/dendritic		

3. Results and Discussion

Through the landscape metrics analysis it was possible to identify some important issues about the theme. Looking to exemplify some points of view, an area was selected aiming to exemplify the consequence of three different cell size (Figure 3). The software used aims to identify some homogeneous characteristics, thus, the size of cell should be adequate to support the machine decision making. Considering that machine decision making is taken according to the examples given by human, the most adequate mean it to provide some very good samples.

According to figure 3-I, it is possible to notice distinct patterns along the cell, as geometric, multidirectional and even diffuse, which could cause some kind of confusion and consequently misclassification. In II, it is possible to notice that a1, a2, b2 and b3 would probably be classified as geometric, c1 as multidirectional, c2, c3 as diffuse, and a3 and b1 the software could present some kind of misclassification. The same method could be applied for III, however, it is possible to notice new others categories, as consolidated (a1, a2, b3, b4, b5, c3, c4, c5) and forest (e6, e4). In none of cases, GeoDMA analyses the context, which can influence the pattern identification and misclassified the data (Figure 4). In this manner, the cell size should be select considering the aim of the study and some characteristics, as accuracy requirement and scale of the map.

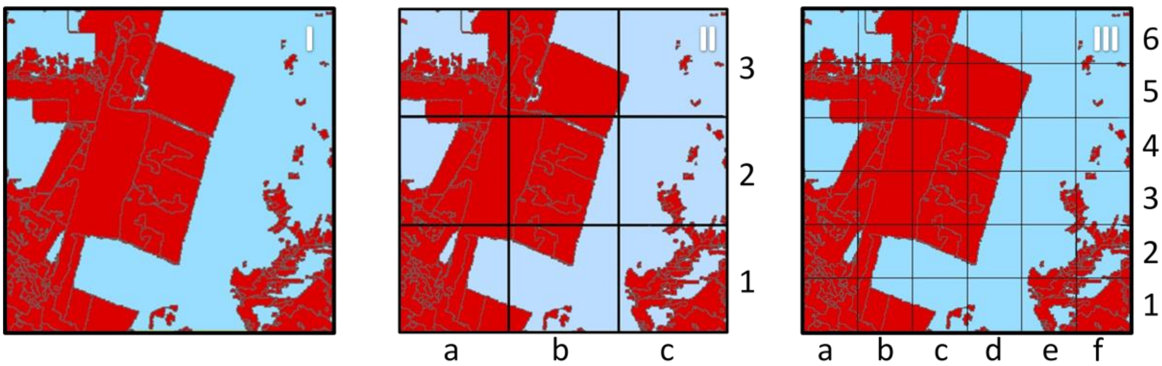


Figure 3. Examples of different landscape metrics. I. cells of 15 km^2; II. cells of 5 km^2; III. cells of 2.5 km^2.

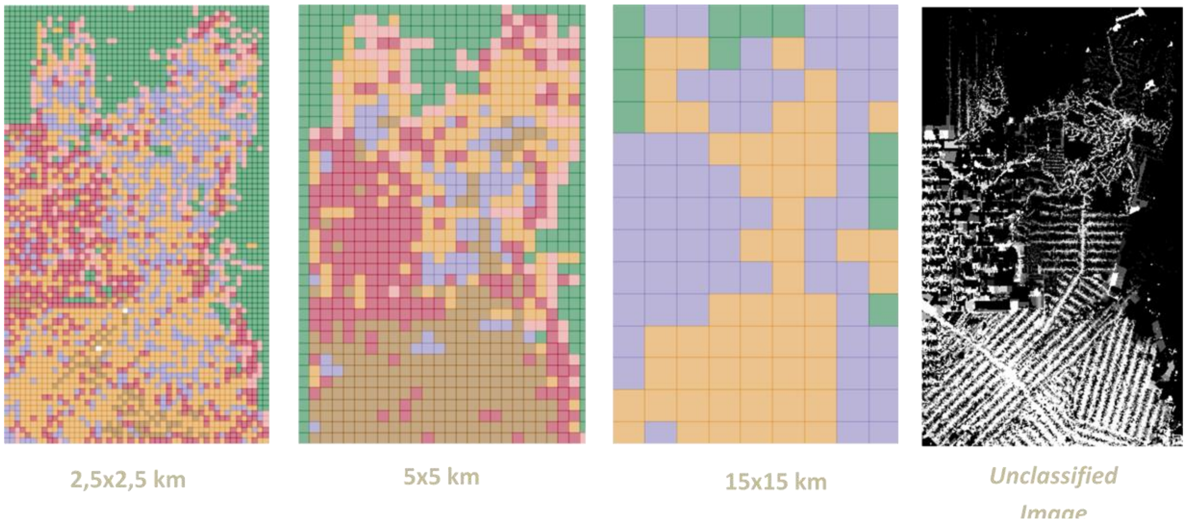


Figure 4. Result of the classification cell grids

For this study, 5 km^2 grid of cells was selected, and the typology table was developed (Table 1). Through the 5 km^2 cells, the final classified image was generated (Figure 5).

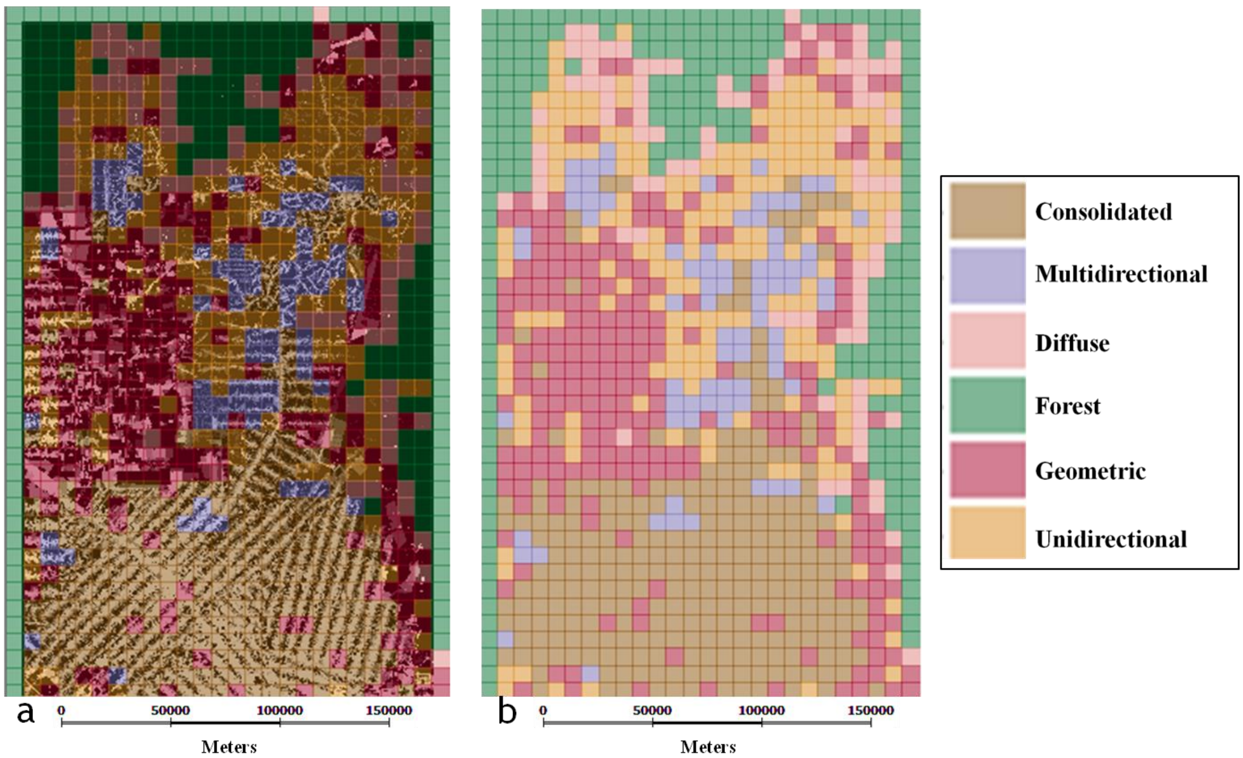


Figure 5. Final classified image. a) classification with raster image; b) final classification.

For that, the decision tree applied for the final classification can be found in figure 6.

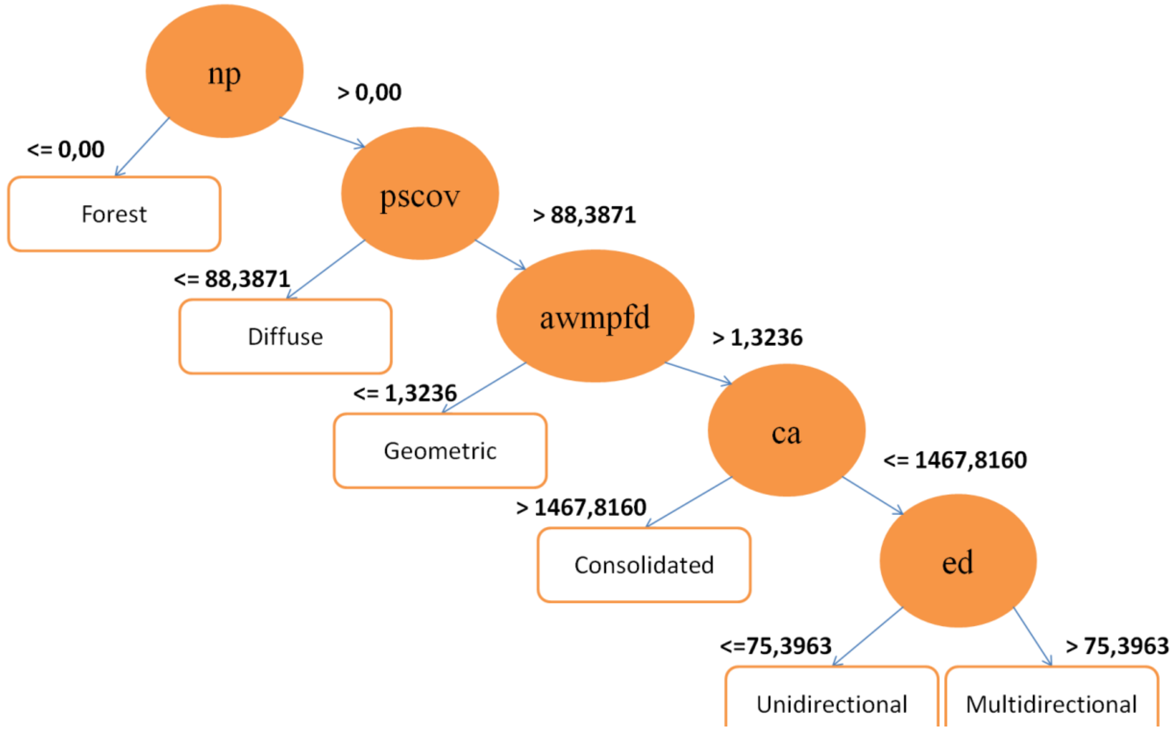


Figure 6. Final decision tree.

Most of the 16 different landscape based features applied by GeoDMA are based on *Fragstats software* (McGarial and Marks 1995). Although, in this decision tree it was identified 5 different features. The description of them can be found in Table 3.

Table 3. Landscape-based features identified in the study area.

Subject area	Name	Description	Formula	Range	Units
Area metrics	ca (class area)	Class Area means the sum of the areas of a cell	$\sum_{j=1}^{n} a_j$	≥ 0	ha
Shape metrics	Awmpfd (area-weighted mean patch fractal dimension)	AWMPFD stands for Area-weighted Mean Patch Fractal Dimension	-	≥ 0	-
Edge metrics	ed (edge density)	Edge Density equals the sum of the lengths (m) of all edge segments involving the corresponding patch type, divided by the total landscape area (m^2)	$\frac{\sum_{j=1}^{m} e_j}{A} 10^{-4}$	≥ 0	m / ha
Patch density, patch size and variability metrics	pscov (patch size coefficient of variation)	Patch Size Coefficient of Variation calculates the ratio between the features _c_pssd_ and _c_mps_	$\frac{\sum_{j=1}^{n} \frac{j}{a_j}}{n}$	≥ 0	m^{-1}
Patch density, patch size and variability metrics	np (number of patches)	NP equals the number of patches inside a particular landscape	n	≥ 0	-

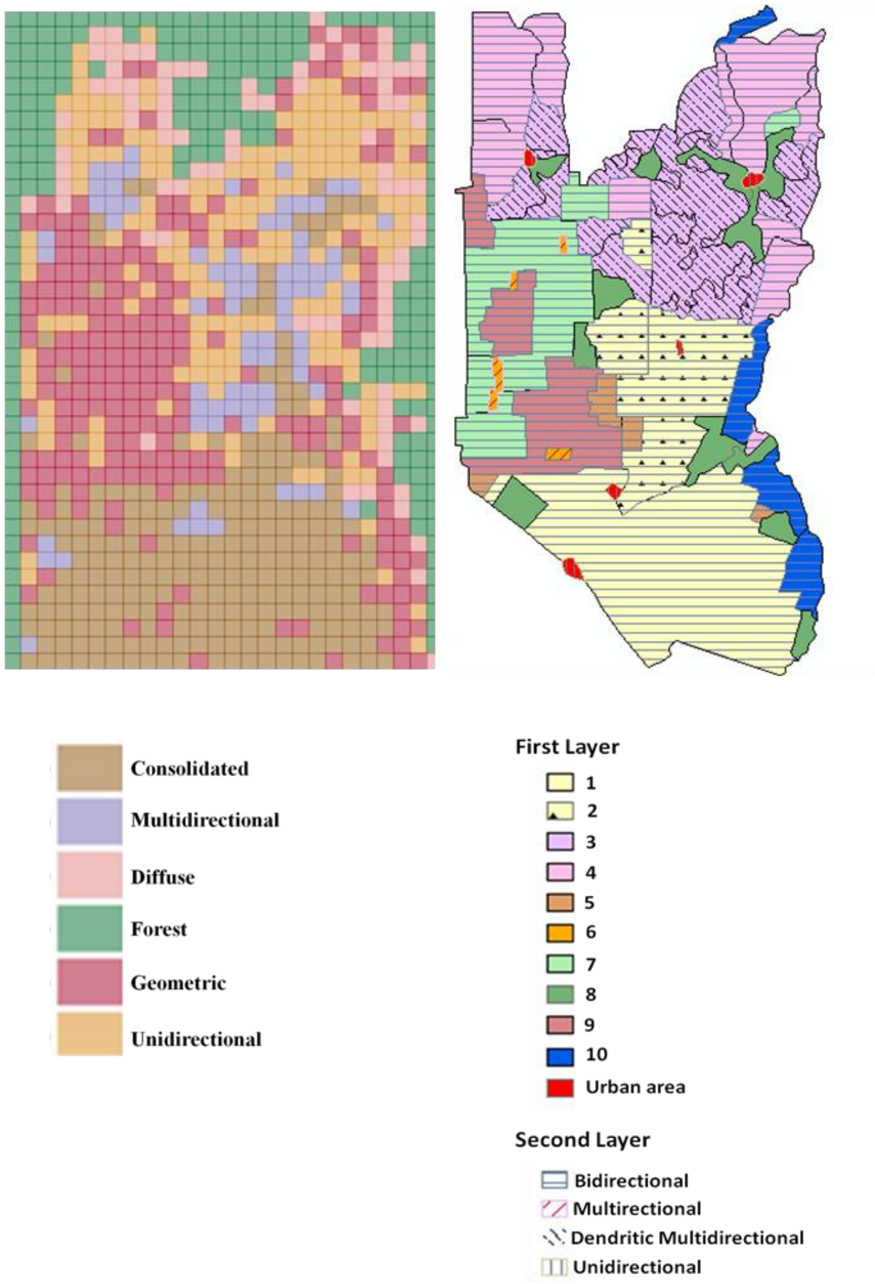


Figure 7. Comparison of maps.

Patterns 1 and 2 were classified as orthogonal/bidirectional by Escada (2003) and roughly GeoDMA was sensible for both patterns, but mainly for the first one, and they are called consolidated (figure 7). For the number 3, the software was not so accurate and there were some misclassifications as unidirectional. In this manner, it is possible to notice that the scale is essential for both pattern. If the scale is too big, it is possible to identify unidirectional, and if it is small, dendritic format. In number 4 and 5, it is not possible to assert conclusion, considering that they were a compilation of different patterns.

Once GeoDMA is applied to identify a considerable proportion of the same pattern, its accuracy was not sensible enough to discriminate pattern 6. The areas 7, 8, 9 and 10 were mostly related to geometric pattern in GeoDMA software. This phenomenon occurred, probably, considering the those areas were classified by Escada (2003) as medium to big plot, which landscape metrics are considered geometric. In other others, there is a high probability that this pattern due to modern machines and big plots of agriculture.

4. Conclusions

Landscape metrics are concepts that can assist the identification of degradation agents process. Furthermore, remote sensing has become an important auxiliary tool in this analysis. However, it is important the continuous development of the metrics as well as remote sensing images and software, considering that the complexity tends to increase along the years. Besides that, tropical forests are located near the equator, where most of the countries are considered undeveloped and because of that, financial resources for this activity can be scare. In this manner, free software as GeoDMA can be a essential tool in order to manage natural resources. In this work, it was possible to notice how important it to apply GeoDMA techniques and metrics to real deforestation patterns, considering that it may support the understanding of influences, as actors and phenomenon. However, some important points must be highlighted: i) GeoDMA is a remarkable tool to earth observation remote sensing, however, there are some error that must be corrected by computer engineers; ii) The size of cell applied must be analyzed in each pattern, that means, that further studies must continue in order to integrate the best landscape metrics for the same area; iii) Machine does not see the context, just the cell, which can drives the work to some false analysis; iv) training phase can take longer, if the samples are not good enough nor much enough; v) if there is not a considerable number of samples, it would not be adequate to apply data mining, but manual classification; vi) the bigger is the size of the cell, the less there are samples to learn machine phase; vii) the study of the area, including historical analysis is an essential step in order to get results closer to reality as possible; viii) the final classification was satisfactory, but even though, more samples and training are necessary aiming better results.

5. References

Aguilera, Francisco, Luis M. Valenzuela, and André Botequilha-Leitão. 2011. “Landscape Metrics in the Analysis of Urban Land Use Patterns: A Case Study in a Spanish Metropolitan Area.” *Landscape and Urban Planning* 99(3):226–38.

Albuquerque e Arraes, Ronaldo de, Francisca Zilania Mariano, and Andrei Gomes Simonassi. 2012. “Causas Do Desmatamento No Brasil E Seu Ordenamento No Contexto Mundial.” *Revista de Economia e Sociologia Rural* 50(1):119–40. Retrieved June 10, 2016 (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-20032012000100007&lng=en&nrm=iso&tlng=pt).

DiBari, John N. 2007. “Evaluation of Five Landscape-Level Metrics for Measuring the Effects of Urbanization on Landscape Structure: The Case of Tucson, Arizona, USA.” *Landscape and Urban Planning* 79(3):308–13.

ESCADA, M. I. S. Evolução de padrões de uso e cobertura da terra na região Centro-Norte de Rondônia. 2003. 264 p. (INPE-10209-TDI/899). Instituto Nacional de

Pesquisas Espaciais, São José dos Campos. 2003. Available at <http://urlib.net/sid.inpe.br/jeferson/2003/06.30.13.31>. Accessed in: jun. 2003.

Fayyad, Usama and Paul Stolorz. 1997. "Data Mining and KDD: Promise and Challenges." *Future Generation Computer Systems* 13(2-3):99–115. Retrieved February 22, 2016 (http://www.sciencedirect.com/science/article/pii/S0167739X97000150).

Geist, Helmut and Eric F. Lambin. 2001. *What Drives Tropical Deforestation? A Meta-Analysis of Proximate and Underlying Causes of Deforestation Based on Subnational Case Study Evidence.* Retrieved (http://www.sciencedirect.com/science/article/pii/095937809090005T).

Haralick, Robert M. and Linda G. Shapiro. 1985. "Image Segmentation Techniques." *Computer Vision, Graphics, and Image Processing* 29(1):100–132.

Husson, A. et al. 1995. *Study of Forest Non-Forest Interface Typology of Fragmentation of Tropical Forest : Catalogue*. TREES. edited by E. Commission. Luxembourg: Luxembourg: Office for Official Publications of the European Communities. Retrieved (http://bookshop.europa.eu/en/study-of-forest-non-forest-interface-pbCLNA16291/;pgid=Iq1Ekni0.1lSR0OOK4MycO9B0000w7D_qKTH;sid=8CkHdWKWuEoHWTXLFh2Z0gCzUup7JcBVjHk=?CatalogCategoryID=r2AKABstX7kAAAEjppEY4e5L).

INPE, Instituto Nacional de Pesquisas Espaciais. PRODES. Available at: <http://www.obt.inpe.br/prodes/index.php>. Accessed in jun. 2016a.

INPE, Instituto Nacional de Pesquisas Espaciais. GeoDMA. Available at: <http://wiki.dpi.inpe.br/doku.php?id=start>. Accessed in jun. 2016b.

INPE, Instituto Nacional de Pesquisas Espaciais. GeoDMA - Geographic Data Mining Analyst. Available in: <http://wiki.dpi.inpe.br/doku.php?id=geodma>. Accessed at jun. 2016c.

Körting, Thales Sehn, Leila Maria Garcia Fonseca, and Gilberto Câmara. 2013. "GeoDMA—Geographic Data Mining Analyst." *Computers & Geosciences* 57:133–45. Retrieved April 28, 2016 (http://www.sciencedirect.com/science/article/pii/S0098300413000538).

Mas, J. F. 1997. "Monitoring Land-Cover Changes: A Comparison of Change Detection Techniques." *International Journal of Remote Sensing* 20(1):139–52.

McGarial, Kevin and Barabara J. Marks. 1995. "FRAGSTAT: Spatial Pattern Analysis Program for Quantifying Landscape Structure." *United States Department of Agriculture, Pacific Northwest Research Station.* (August):120 pages.

McGarigal, Kevin. 2013. "Landscape Pattern Metrics." in *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd. Retrieved (http://dx.doi.org/10.1002/9780470057339.val006.pub2).

Mertens, Benoît and Eric F. Lambin. 1997. "Spatial Modelling of Deforestation in Southern Cameroon." *Applied Geography* 17(2):143–62. Retrieved June 12, 2016 (http://linkinghub.elsevier.com/retrieve/pii/S0143622897000325).

Rogan, John and Dongmei Chen. 2004. "Remote Sensing for Mapping and Monitoring

Land-Cover and Land-Use Change - an Introduction." *Progress in Planning* 61(4):301–25. Retrieved (http://gis.geog.queensu.ca/publicationpdfs/2004rsinplanning.pdf).

Silva, Marcelino Pereira dos Santos, Gilberto Câmara, Ricardo Cartaxo Modesto Souza, Dalton Morrison Valeriano, and Maria Isabel Sobral Escada. 2005. "Mining Patterns of Change in Remote Sensing Image Databases." *IEEE International Conference on Data Mining (ICDM'05)* 8. Retrieved (http://urlib.net/dpi.inpe.br/hermes2@1905/2005/10.03.07.12).

Soares-Filho, Britaldo Silveira et al. 2005. "Cenários de Desmatamento Para a Amazônia." *Estudos Avançados* 19(54):137–52. Retrieved June 10, 2016 (http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142005000200008&lng=en&nrm=iso&tlng=pt).

Trani, Margaret Katherine and Robert H. Giles, Jr. 1999. "An Analysis of Deforestation: Metrics Used to Describe Pattern Change." *Forest Ecology and Management* 114(2):459–70.