



UTILIZAÇÃO DE MINERAÇÃO DE DADOS E MÉTRICAS DE PAISAGEM PARA O MAPEAMENTO AUTOMATIZADO DE TIPOLOGIAS DE PAISAGEM NA AMAZÔNIA ORIENTAL

R.R. S.Oliveira¹, A. Venturieri²

¹Universidade do Estado do Pará, Brasil

²Empresa Brasileira de Pesquisa Agropecuária, Brasil

Comissão VI - Sistemas de Informações Geográficas e Infraestrutura de Dados Espaciais

RESUMO

Este artigo tem como objetivo apresentar o mapeamento de tipologias de paisagem através de mineração de dados e métricas de Ecologia de Paisagem, utilizando os dados do projeto TerraClass Amazônia para os anos de 2008 e 2010. Como recorte foram selecionadas as Regiões de Integração do Araguaia e Tapajós, no Estado do Pará, devido suas distintas e intensas dinâmicas de uso. Neste sentido, a Mineração de Dados foi proposta, se constituindo como uma das tecnologias mais promissoras na identificação e análise de dados, transformando-os em informações. De acordo com as análises realizadas, foram identificadas sete tipologias de paisagem. A RI do Tapajós, apresentou mais áreas classificadas como Paisagens Florestais, tanto para o ano de 2008 quanto para 2010. Já a RI do Araguaia, apresenta maior predomínio de Tipologias de Paisagem consolidadas, mais especificamente a TP com Pecuária Consolidada. Portanto, o mapeamento automatizado de Tipologias de Paisagem utilizando o *Plugin GeoDMA* do Software *Terra View* demonstrou-se eficaz e preciso, visto que os resultados alcançados apresentam coerência com a realidade.

Palavras-chave: Mineração de Dados, Ecologia de Paisagem, GeoDMA, Mudança de Uso da Terra, Amazônia.

ABSTRACT

This article aims to present the mapping of landscape typologies through data mining and landscape ecology metrics, using data from the TerraClass Amazonia project for the years 2008 and 2010. As a cut, the Integration Regions of Araguaia and Tapajós, in the State of Pará, due to its different and intense dynamics of use. In this sense, Data Mining was proposed, constituting one of the most promising technologies in the identification and analysis of data, transforming them into information. According to the analysis, seven types of landscape were identified. The IR of the Tapajós, presented more areas classified as "Forest Landscapes", both for the year of 2008 and for the year of 2010. The RI of the Araguaia, presents a greater predominance of Consolidated Landscape Typologies, more specifically the TP with Consolidated Livestock. Therefore, the automated mapping of Landscape Typologies using the Terra View GeoDMA Plugin has proven to be effective and accurate, as the results achieved are consistent with reality.

Keywords: Data Mining, Landscape Ecology, GeoDMA, Land Use Change, Amazônia.

1- INTRODUÇÃO

A significativa evolução computacional durante as últimas décadas e a disseminação das geotecnologias têm dado subsídio a implementação de Banco de Dados Geográficos cada vez mais robustos, com isso o volume na geração de dados (espaciais e não-espaciais) é cada vez maior, bem como os meios de consulta a eles. No entanto, torna-se uma necessidade buscar meios e técnicas eficazes de análise desses dados armazenados.

Para tanto, no final da década de 80, a Mineração de Dados (Data Mining, em inglês) foi proposta, se constituindo atualmente como uma das tecnologias mais promissoras na identificação e análise de dados, transformando-os em informações

(conhecimento). Segundo Fayyad (1997), no modelo tradicional para transformação dos dados em informação, consiste em um processamento manual de todas essas informações por especialistas. E, devido ao grande volume de dados, se torna inviável.

Atualmente, são vastas as referências que definem e padronizam as fases da Mineração de Dados ou KDD (Knowledge Discovery in Databases ou Descoberta de Conhecimento nas Bases de Dados), no entanto, todos apresentam basicamente a mesma estrutura (Hand *et al.*, 2008; Larose, 2005). Ressalta-se que, há discordância entre os autores se Mineração de Dados e KDD podem ser considerados sinônimos ou se a Mineração de Dados é apenas uma das etapas do processo de KDD.

De acordo com Goldschmidt (2004), existem diferentes métodos de Mineração de Dados (*Data mining*), entre eles destacam-se: métodos baseados em redes neurais, métodos baseados em algoritmos genéticos, métodos baseados em instâncias, métodos estatísticos, métodos específicos, métodos baseados em indução de árvores de decisão e métodos baseados em lógica nebulosa.

A mineração de dados geográficos envolve a aplicação de ferramentas computacionais para revelar padrões em objetos e eventos distribuídos no espaço geográfico e ao longo do tempo. Estes padrões podem envolver as propriedades espaciais de objetos individuais e eventos (a forma, a extensão, etc) e as relações espaço-temporais entre os objetos e eventos, além dos atributos não-espaciais de interesse na mineração de dados tradicional (Hand *et al.*, 2008).

Existem disponíveis uma diversidade de softwares e algoritmos voltados à mineração de dados. Todavia, nesta pesquisa, optou-se pelo *plugin GeoDMA* (INPE, 2012) do *software* livre *Terra View* (INPE, 2012) e pelo algoritmo C4.5 (Quinlan, 1993) que faz uso da técnica de mineração de dados pela indução de árvores de decisão. Segundo Vieira (2010), o C4.5 possui uma série de parâmetros a saber: 1) fator de confiança; 2) número mínimo de instâncias por nó-folha; 3) número de ramificações; 4) poda ou não-poda, que podem ser configurados. Evidencia-se que os ajustes nestes parâmetros podem auxiliar na melhoria do resultado final de uma classificação.

O *GeoDMA* é um *software* livre, sendo desenvolvido em C++ e, *Quad Tree* (QT) para a construção da interface com o usuário. Utiliza biblioteca *TerraLib* visando as operações geográficas e de processamento de imagens. O *GeoDMA* é um sistema de análise de imagem que integra ferramentas de: análise de imagem, métricas com base na teoria da Ecologia de Paisagem, tratamento de características multi-temporais, e as técnicas de mineração de dados. O sistema baseia-se na metodologia proposta por Silva *et al.* (2005), para identificar padrões de desmatamento na Amazônia. É uma solução de *software* livre para aplicações em sensoriamento remoto, rodando em diferentes plataformas, por exemplo, *Windows* e *Linux* (Körting *et al.*, 2013).

O método de mineração proposto por Silva *et al.* (2005) foi desenvolvido para uma aplicação específica de identificação de padrões de desmatamento. Para o desenvolvimento da metodologia, os seguintes passos foram implementados: a) definição de uma tipologia de padrões, de acordo com a aplicação desejada; b) construção de um conjunto de referência com padrões espaciais; c) mineração do banco de dados por meio de um classificador estrutural.

Assim, características são extraídas. Os atributos espaciais incluem área, perímetro, dimensão fractal, grau de retangularidade e ângulo

principal, que são calculados e armazenados nas tabelas de referência do sistema *Terra View*; Visualização dos atributos: dado o conjunto de atributos, o usuário possui uma ferramenta muito importante para realizar a análise exploratória, através da visualização do espaço de atributos. Dois atributos são selecionados e um gráfico de dispersão mostra a separabilidade dos dados.

2- RECORTE ESPACIAL DE ANÁLISE

A pesquisa tem como foco de análise as Regiões de Integração do Tapajós e Araguaia no Estado do Pará. Estas áreas foram escolhidas por estarem sob influências de dois eixos rodoviários que desde suas implantações têm grande destaque no contexto do histórico de ocupação da Amazônia Legal.

A Região de Integração do Araguaia está localizada no Sudeste do Estado do Pará, engloba quinze municípios do Estado do Pará, sendo eles: Água Azul do Norte; Bannach; Conceição do Araguaia; Cumaru do Norte; Floresta do Araguaia; Ourilândia do Norte; Pau d'Arco; Redenção; Rio Maria; Santa Maria das Barreiras; Santana do Araguaia; São Félix do Xingu; Sapucaia; Tucumã e Xinguara (IDESP, 2013). Segundo os dados do IDESP (2013) Abrangendo uma área territorial de 174.140 Km² (IDESP, 2013), e uma população de 472.930 habitantes (IBGE, 2011), com densidade demográfica de 2,72 (Hab./Km²).

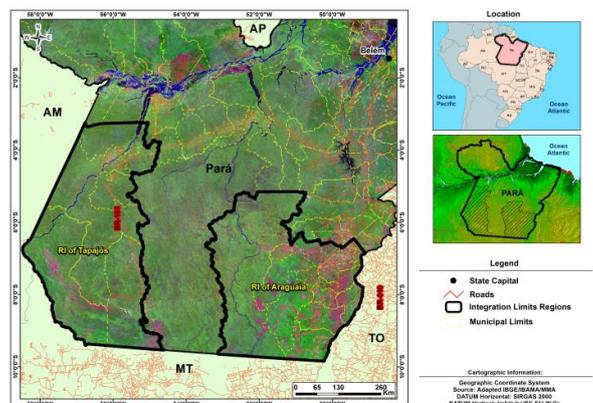


Fig. 1: Mapa de localização das RI.

Já a RI do Tapajós está localizada no Sudoeste do Estado do Pará, engloba os municípios de Aveiro; Itaituba; Jacareacanga; Novo Progresso; Rurópolis e Trairão (IDESP, 2013). Abrange uma área territorial de 189.610 Km², com uma população de 209.530 habitantes (IBGE, 2011), com densidade demográfica de 1,11 Hab./Km² (IDESP, 2013).

3- ELABORAÇÃO DO BANCO DE DADOS

Para a realização dos processamentos foram utilizados os dados de Uso e Cobertura da Terra da Amazônia Legal elaborado no âmbito do projeto TerraClass (2008 e 2010) do Instituto Nacional de

Pesquisas Espaciais (INPE). Assim como as bases cartográficas contendo: limites municipais, sedes municipais, hidrografia, limites das Regiões de Integração, estradas, dentre outras disponíveis no banco de dados georreferenciados do Instituto Brasileiro de Geografia e Estatística (IBGE).

Portanto, O banco de dados georreferenciados foi estruturado na plataforma do *software Terra View 4.0.0* por se constituir a versão mais estável que possui compatibilidade com o plugin *GeoDMA 2.1* sendo utilizado para a mineração dos dados.

Posteriormente, considerando uma malha espacial celular, foi realizada uma análise de sensibilidade para escolha do tamanho do box de células. O objetivo desta análise foi definir a extensão da área de uma unidade de análise/paisagem, para a construção da tipologia de padrões de ocupação a ser utilizada na fase de mineração e classificação dos padrões nas áreas de estudo. Com células de 196 km².

A escolha das métricas a serem utilizadas durante o processo de classificação foi realizada com base na frequência das métricas apresentadas nas árvores de decisão durante o processo de testes preliminares realizados em laboratório. Amostras de treinamento foram selecionadas no conjunto de dados dos anos analisados [2008 e 2010].

Como resultado, obtém-se uma árvore de decisão com as métricas que melhor discriminaram as feições, sendo gerados mapas de padrões de ocupação para cada ano. Esses mapas podem ser analisados separadamente e/ou podem ser combinados para gerar um mapa com as principais trajetórias de ocupação.

Visando a validação dos resultados obtidos através de processos automatizados em Sistemas de Informações Geográficas (SIG's), foram realizados sobrevoos com helicóptero sobre pontos aleatórios, sorteados através de operações estatísticas no *software Terra View 4.0.0*.

4- AS TIPOLOGIAS DE PAISAGEM DAS REGIÕES DE INTEGRAÇÃO DO ARAGUIA E TAPAJÓS

De acordo com as análises realizadas, foram identificadas sete Tipologias de Paisagem - TP. A nomenclatura escolhida para cada tipologia de paisagem foi adotada visando à composição e predominância de uso e cobertura da terra de cada uma, que identifica um estágio de ocupação e tipo de uso. Assim, obteve-se o seguinte modelo conceitual:

A Paisagem Florestal (PF) é caracterizada pelo predomínio de áreas florestadas com boa integridade em sua estrutura, árvores de médio a alto porte, diversidade de espécies de fauna e flora, sem significativas alterações por atividades antrópicas.

A Paisagem com Agricultura Familiar em Estágio Inicial (PAFEI) compreendem as áreas em

estágio inicial de ocupação (Becker, 2001; Ferreira, 2001 e Venturieri 2003), caracterizadas por pequenos estabelecimentos rurais, dispersos e fragmentados, sem títulos de terra, geralmente concentrados ao longo das rodovias e suas vicinais.

A Paisagem com Agricultura Familiar Estabelecida (PAFE) corresponde as áreas em estágio intermediário de ocupação, apresentando pequenos e médios estabelecimentos rurais, com uma maior fixação do homem à terra e um maior emprego de infraestruturas às propriedades rurais, apresentando um nível intermediário de antropização da terra.

A Paisagem com Agricultura Consolidada (PAC) representa as áreas em avançado estágio de ocupação, com grandes estabelecimentos rurais, geralmente com monocultura empresarial de grãos, com elevado grau de mecanização, financiamento agrícolas, apresentando um nível elevado de antropização da terra pelo homem e supressão dos remanescentes florestais.

A Paisagem com Pecuária Consolidada (PPC) engloba as áreas em avançado estágio de ocupação, com grandes estabelecimentos rurais com atividades econômicas direcionadas à pecuária extensiva leiteira ou de corte, com manejo de pastagens mecanizado, apresentando baixos índices de remanescentes florestais e alto grau de antropização.

A Paisagem Urbana (PU) representa as áreas em estágio avançado de ocupação, com a presença de conglomerados urbanos, representando pequenas e médias cidades, bem como os grandes centros urbanos.

E, a Paisagem Ribeirinha (PR) corresponde às áreas em estágio inicial de ocupação, com pequenos e médios estabelecimentos rurais localizados às margens dos rios, são intituladas popularmente como "ocupações ribeirinhas" (Trindade Júnior, 2008).

A RI do Tapajós (Fig. 2 e 3), assim como suposto, apresentou mais áreas classificadas como Paisagens Florestais, tanto para o ano de 2008 quanto para o ano de 2010. Pode-se verificar que essas paisagens estão, em sua maioria, localizadas em Unidades de Conservação, sendo resguardadas por lei, fato que minimiza a pressão da fronteira agrícola sobre os remanescentes florestais que ainda restam na região.

Resolvendo que, mesmo nessas áreas e ainda com as restrições legais estabelecidas em lei, podem-se encontrar modificações nessas paisagens dentro das UC's e Terras Indígenas, com a modificação da paisagem principalmente por pequenos produtores rurais que se instalam nessas áreas alterando a sua estrutura florestal original.

Assim, pode-se verificar modificação em aproximadamente 31,34% das tipologias de paisagem florestal para as demais tipologias de paisagem, principalmente para a TP Agricultura Familiar em Estágio Inicial, com conversão direta de 30,06%.

É importante ressaltar que houveram algumas confusões com o classificador, visto que o processo foi automatizado, tal comportamento pode ser observado em algumas conversões inconsistentes, isto correspondeu a 9,23% do total de amostras. No entanto, essa estimativa de erro pode ser totalmente corrigida através de edição manual do operador.

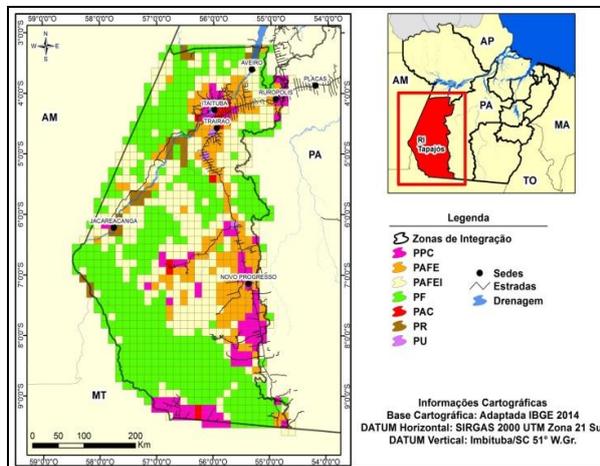


Fig. 2: Mapa de Tipologia de Paisagem da Região de Integração do Tapajós, ano 2008.

Analisando o resultado da classificação de TP realizada para a RI do Tapajós em ambos os anos pode-se constatar que a “evolução” de uma paisagem mais íntegra (florestada) para uma região mais antropizada está ocorrendo ao longo da BR-163 e suas vicinais, portanto, infere-se que a mesma possui forte influência para a modificação do seu entorno.

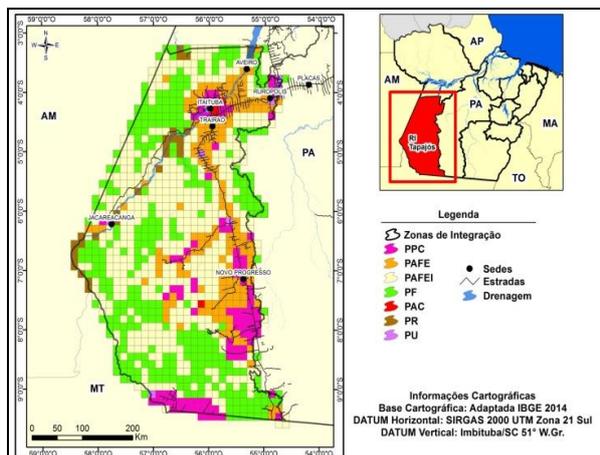


Fig. 3: Mapa de Tipologia de Paisagem da Região de Integração do Tapajós, ano 2010.

Já na RI do Araguaia (Fig. 4 e 5) ao contrário do que foi verificado para a RI do Tapajós, apresenta maior predomínio de Tipologias de Paisagem consolidadas, mais especificamente a TP com Pecuária Consolidada, cerca de 41,54% da paisagem da região.

Entretanto, nota-se que a região também apresenta em sua estrutura TP's com remanescentes

florestais, que correspondem a 27,12% da paisagem, tal como verificado à RI do Tapajós estas TP's também estão concentradas em Unidades de Conservação, mas principalmente em Terras Indígenas.

Sendo que, do ano de 2008 para o ano de 2010, há uma perda de paisagens florestais, verificando-se conversões diretas para TP's com Agricultura Familiar em Estágio Inicial (4,5%). Indicando assim que mesmo se constituindo enquanto uma paisagem predominantemente “consolidada” ainda há uma inserção de áreas florestadas ao processo produtivo, no entanto, essa taxa representa menos que da metade dos valores encontrados para essa mesma conversão na RI do Tapajós, inferindo assim, que a Região do Tapajós apresenta maior dinâmica da paisagem se comparada a RI do Araguaia.

Ressalta-se que, grande parte dessas Conversões PF para PAFEI estão localizadas em Terras Indígenas, sendo verificadas intensas alterações na Terra Indígena dos Kayapós, ou seja, mesmo sendo resguardadas por legislação específica, verifica-se modificações dessas áreas florestadas, com a implantação de pastagem cultivada, por médios e grandes pecuaristas.

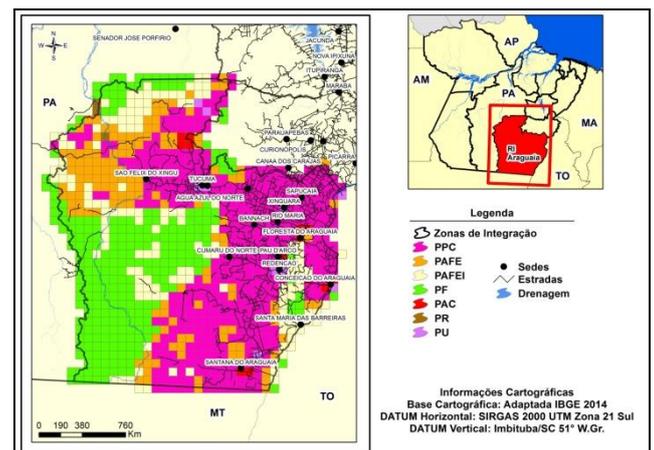


Fig. 4: Mapa de Tipologia de Paisagem da Região de Integração do Araguaia, ano 2008.

No que tange as TP's antropizadas, verifica-se que as mesmas, também, “evoluem” em estágio de antropização à medida que se aproximam da BR-010 e suas vicinais. Sendo que, tanto para o ano de 2008 quanto para o ano de 2010 as Tipologias de Paisagem Consolidadas, ou seja, que tiveram sua estrutura original totalmente alteradas, podem ser identificadas ao longo das margens da BR e suas vicinais.

Já as TP's com Agricultura Familiar Estabelecida, ao contrário do que foi verificado para a RI do Tapajós, não possuem tanta expressividade para a RI do Araguaia, visto que a mesma representa um estágio intermediário entre a implantação de um sistema produtivo e sua consolidação na paisagem, posto que a paisagem desta região é dominada por médios e grandes latifúndios pecuaristas.

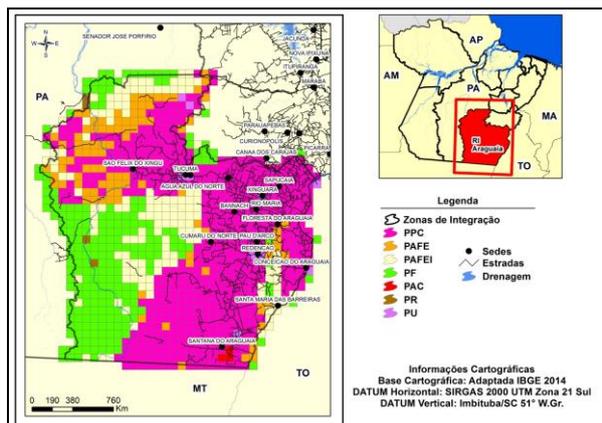


Fig. 5: Mapa de Tipologia de Paisagem da Região de Integração do Araguaia, ano 2010.

Outro fato relevante se refere a maior concentração de Paisagens Urbanas na RI do Araguaia, se comparada a RI do Tapajós.

5- CONSIDERAÇÕES FINAIS:

A Região de Integração do Araguaia apresenta tipologias de paisagens que possuem uso consolidado, estando em grande parte associadas a extensas áreas de pastagens cultivadas com pecuária extensiva e de corte, diferente das tipologias de paisagens da RI do Tapajós que se caracterizam por apresentar predominância das tipologias de paisagem com Agricultura Familiar em Estágios Iniciais de ocupação, associadas às pequenas e médias propriedades rurais.

O processo de mapeamento automatizado de Tipologias de Paisagem utilizando o *Plugin GeoDMA* do *Terra View* demonstrou-se eficaz e preciso, visto que os resultados alcançados apresentam coerência com a realidade de cada Região de Integração. Fato este constatado após se confrontar as bases cartográficas de Instituições de pesquisa científica confiáveis, com os mapas de Tipologias de Paisagens gerados pelo *software*.

O confronto entre os dados confirmaram que as TP com as classes que representam os estágios de ocupação de uma paisagem, encontram-se em áreas com malha fundiária registrada no Cadastro Ambiental Rural – CAR, da Secretaria de Estado de Meio Ambiente do Estado do Pará e que as áreas com cobertura vegetal mais preservada encontram-se em áreas protegidas pela legislação ambiental.

REFERÊNCIAS BIBLIOGRÁFICAS

Becker, B. K. Revisão das políticas de ocupação da Amazônia: é possível identificar modelos para projetar cenários? *Parcerias Estratégicas*, v. 6, n. 12, p. 135-158, 2001.

Fayyad, U., Stolorz, P. Data mining and KDD: promise and challenges. *Future Generation Computer Systems* 13, 99-115, 1997.

Ferreira, L. A. Le role de élevage bovin dans La viabilité agroécologique et socio-économique des systèmes de production agricoles familiaux en Amazonie brésilienne - Le cas d'Uruará (Pará, Brésil). *Institut National Agronomique Paris-Grignon*, Paris, 2001.

Goodchild, M. GIScience, geography, form, and process. In: Association of American Geographers. Vol. 94. *Blackwell Publishing*, Oxford, UK, pp. 709-714., 2004.

Han, J. e Kamber, M.. Data Mining: Concepts and Techniques. Tech. rep., *University of Illinois at Urbana-Champaign*. 2008. URL <http://books.google.com/books?id=AfL0t-YzOrEC>

IBGE. Instituto Brasileiro de Geografia e Estatística. Sinopse do Senso Demográfico de 2010. Rio de Janeiro, 2011.

IDESPE. Instituto de Desenvolvimento Econômico Social e Ambiental do Pará. Perfil da gestão ambiental dos municípios paraenses: programa municípios verdes. Belém: IDESP, 2013.

INPE. Instituto de Pesquisas Espaciais. 2012. TerraView. URL <http://www.dpi.inpe.br/terraview>

INPE. Instituto de Pesquisas Espaciais. GeoDMA, Geographical Data Mining Analyst, 2007. Disponível em:

<<http://www.dpi.inpe.br/geodma/?lingua=portugues> >. Acesso em: Setembro 2012.

Korting, T. S.; Fonseca, S.M.G.; Câmara, G. GeoDMA – Geographic Data Mining Analyst. *Computers & Geosciences*, vol. 57, 133-145, August, 2013.

Larose, D. T. Discovering Knowledge in *Data: An introduction on Data Mining*. Book. 27-65, 2005.

Quinlan, J. C4.5: programs for machine learning. *Morgan Kaufmann*, San Mateo, CA, 1993.

Silva, M. P. S.; Câmara, G.; Escada, M. I. S.; Souza, R. C. M.; Valeriano, D. M. Mining patterns of change in remote sensing image databases. *15 IEEE International Conference on data mining*, 2005.

Trindade Júnior, Saint-Clair Cordeiro da; Tavares, Maria Goretti da Costa. (orgs). *Cidades Ribeirinhas da Amazônia: mudanças e permanências*. Belém: *EDUFPA*, 2008.

Venturieri, A. A dinâmica da ocupação pioneira na rodovia Transamazônica: uma abordagem de modelos de paisagem. 2003, 167f. (*Tese Doutorado*) Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Campus Rio Claro, 2003.

Vieira, M. A. Análise de imagem orientada a objeto e mineração de dados aplicadas ao mapeamento da cultura da cana-de-açúcar, (*Diss. Mestrado*). Instituto Nacional de Pesquisas Espaciais, 2010.