

Classificação de áreas alagáveis da planície do rio Amazonas utilizando mineração de dados e GEOBIA

Felipe Augusto Ventura da Silva Alfaya¹

Mariane Souza Reis¹

Teresa Gallotti Florenzano¹

Cláudio Clamente Faria Barbosa²

¹ Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Sensoriamento Remoto
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{alfaya, reis, teresa}@dsr.inpe.br

² Instituto Nacional de Pesquisas Espaciais – INPE
Divisão de Processamento de Imagens
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
claudio@dpi.inpe.br

Abstract. The study of the Amazon River floodplain is of great importance for many subjects. The accurate definition of its area is the first challenge for this study. An important source of data for this is the SRTM Digital Elevation Model. The objective of this work is to evaluate the use of data mining procedures to map wetlands in the Amazon River floodplain, using SRTM data and HAND. The results obtained from data mining were compared with classifications created using manual selection of attributes. To apply the method two reaches of this floodplain were selected: Codajás, localized in Amazonas State, and Óbidos, localized in Pará State. The classifications were carried out using two segmentation levels and decision trees built using the J48 algorithm implemented in WEKA. The layers of SRTM-DEM, HAND-DEM and HAND-DEM derived slope and curvature images were used in the classification. A Monte Carlo method based analysis was used to evaluate the level of agreement between the classifications and a reference map. Those results were compared to maps created with manual selection of attributes. For the Óbidos reach, both manual selection and data mining yielded similar results but for the Codajás reach the data mining method performed noticeably better as indicated by the Monte Carlo analysis.

Palavras-chave: Amazon River floodplain, data mining, SRTM, HAND, planície de inundação do rio Amazonas, mineração de dados, SRTM, HAND.

1. Introdução

Um mapa da planície de inundação do rio Amazonas é importante tanto para estudos de geomorfologia, quanto para aqueles relativos ao desflorestamento (RENÓ et al., 2011), circulação da água (BARBOSA, 2005; JUNK, 1996), produção pesqueira e qualidade da água (JUNK, 1996), entre outros. A elaboração desse tipo de mapa, no entanto, não é simples, principalmente devido à complexidade dessa área.

No mapeamento de áreas alagáveis, podem ser utilizados dados de sensores ópticos (MERTES et al., 1995; SHIMABUKURO et al., 2002), sensores ativos como o radar (HESS et al., 2003) e/ou o Modelo Digital de Elevação (MDE), derivado de dados ópticos (por ex. ASTER-TERRA) ou de sistemas SAR, tais como o SRTM (RABUS et al., 2003). Os dados topográficos obtidos do SRTM, disponíveis gratuitamente para quase todo o planeta, tem sido utilizados em vários estudos geomorfológicos. No projeto Topodata (VALERIANO e ROSSETTI, 2012) esses dados foram refinados de

3'' (aproximadamente 90 m de resolução espacial) para 1'' (cerca de 30 m) e a partir dos dados refinados foram derivados dados geomorfométricos como, por exemplo, declividade, curvatura horizontal e vertical (VALERIANO E ROSSETTI, 2012). Também a partir dos dados SRTM, Rennó et al. (2008) propuseram o descritor HAND, que gera um MDE tal que rios e lagos possuem altura zero, normalizando o terreno em função da rede de drenagem.

Assim, o conjunto de variáveis, as quais são formadas pelos dados coletados, e de atributos que podem ser derivados desses dados, disponíveis para mapear áreas alagáveis é variado. Em termos de classificação, é comum pensar que um maior número de variáveis para representar uma classe aumenta o poder discriminatório do classificador, o que nem sempre é verdade. Pode ocorrer uma degradação da exatidão dos resultados de classificação com o aumento da dimensionalidade dos dados, ao se manter constante o número de amostras, fenômeno conhecido como a maldição da dimensionalidade (JAIN et al., 2000; OLIVEIRA et al., 2005).

Segundo Tan et al. (2009), pode-se reduzir a dimensionalidade dos dados com o uso de técnicas de álgebra linear, ou com a seleção de um subconjunto de variáveis, eliminando informações redundantes e irrelevantes. Esta seleção pode ser feita de forma manual, utilizando o conhecimento do analista, ou baseada em distâncias estatísticas, ou ainda com o uso de um minerador de dados, que extrai informações úteis da busca realizada em bancos de dados.

Dentre os *softwares* dedicados à mineração de dados, destaca-se o WEKA (Waikato Environment for Knowledge Analysis) (WITTEN et al., 1999). Neste *software*, além do conjunto de algoritmos para tarefas de mineração de dados, existem ferramentas para o pré-processamento dos dados, classificação, regressão, agrupamento, regras de associação e visualização.

A partir da mineração de dados, é possível implementar os resultados diretamente na construção de classificações, ou combiná-los com outras técnicas como, por exemplo, a GEOBIA (*Geographic Object Image Analysis*).

O objetivo deste trabalho foi avaliar a viabilidade do uso de GEOBIA com atributos gerados automaticamente por mineração de dados, utilizando o WEKA, na classificação de áreas de inundação de dois setores da planície Amazônica, Codajás e Óbidos, com diferentes características morfológicas. Os mapas gerados por este processo foram comparados àqueles obtidos por Alfaya (2012), que utilizou um subconjunto de atributos selecionados manualmente.

2. Área de Estudo

O mapa de localização das áreas teste Codajás e Óbidos, é apresentado na Figura 1.

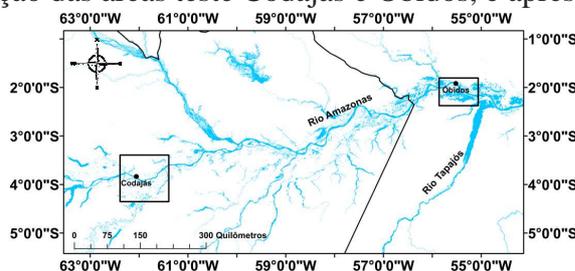


Figura 1. Mapa de localização das áreas de estudo de Codajás e Óbidos.

A planície de inundação de Codajás, de acordo com Mertes et al (1996), caracteriza-se morfológicamente como uma área de transição entre os setores à jusante e à montante do Solimões/Amazonas. Nesta região, o rio é menos meandrante, em relação aos setores à jusante e as taxas de migração de canal são menores. Por outro lado, ocorre

um grande número de pequenos tributários, os quais dão uma forma mais complexa à planície de inundação quando comparada aos setores da montante. Além disso, a constante deposição de sedimentos e a diminuição do gradiente do rio formam uma transição gradual entre a planície de inundação e as áreas não inundáveis (chamadas de terra firme). Isso é mais notável na margem esquerda do rio.

De acordo com Mertes et al (1996), a planície de inundação na área de Óbidos é mais confinada e retilínea, com a presença de lagos de grande extensão. Nesta região, a diferença de altura entre a planície de inundação e as áreas de terra firme é significativa, sendo aproximadamente de 10 m em média. Esta área também apresenta poucos afluentes de pequeno porte, o que facilita o processo de classificação.

3. Material e Métodos

O mapa de referência usado neste trabalho para o cálculo das estatísticas de validação é baseado no elaborado por Hess et al. (2003), combinado com a rede de drenagem extraída do SWBD (*SRTM Water Body Dataset*), e mostrado na Figura 2. Além desse mapa foi utilizado também aquele obtido por Alfaya (2012) no experimento de adequação da árvore de classificação para cada área de estudo, com aplicação de GEOBIA e seleção manual de atributos (Figura 3).

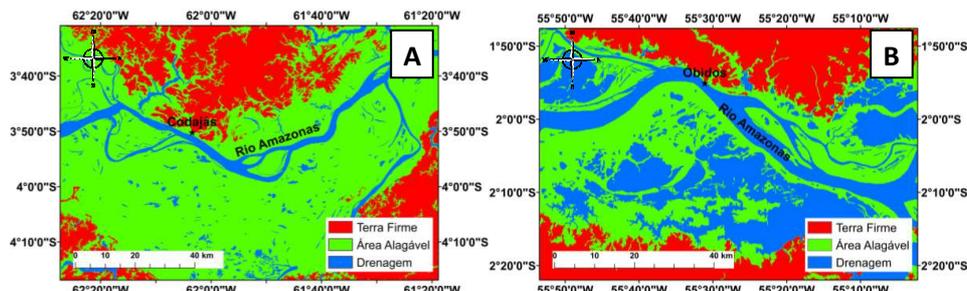


Figura 2. Mapas das áreas de estudo de Hess et al. (2003), A: Codajás. B: Óbidos.

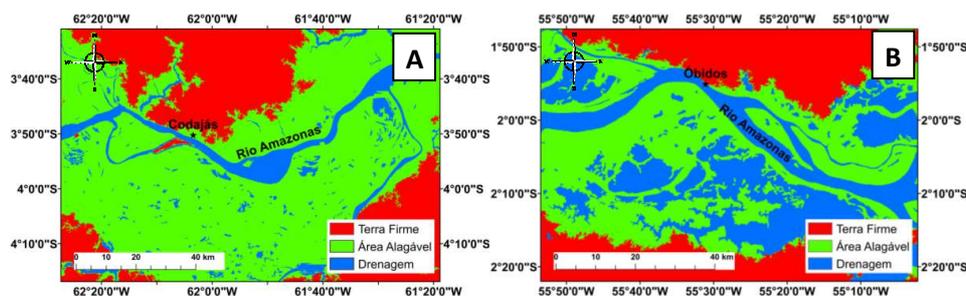


Figura 3. Mapas resultantes da aplicação de GEOBIA com base na seleção manual dos atributos das áreas de estudo, A: Codajás. B: Óbidos. Fonte: Alfaya (2012).

A classificação baseada em objeto e com mineração de dados foi realizada aplicando os seguintes procedimentos: geração do MDE-HAND e das grades de declividade e curvatura horizontal, baseadas neste MDE; geração da árvore de classificação por mineração de dados; mapeamento das áreas alagáveis; validação da classificação.

3.1. Geração do MDE-HAND e das grades declividade e curvatura

Um MDE-HAND para cada área de estudo foi gerado de acordo com os procedimentos descritos em Rennó et al. (2008), a partir do MDE-SRTM. Foi usado um

valor de 10.000 células para o parâmetro de área de contribuição mínima. A partir do MDE-HAND, foram geradas grades de declividade e de curvatura horizontal da área de estudo. O MDE-SRTM, o MDE-HAND e as grades de declividade e curvatura horizontal formam o conjunto de dados utilizados no processo de classificação. Esse foi o mesmo conjunto utilizado para a construção dos mapas das áreas alagáveis em Alfaya (2012), e foi mantido para este trabalho para permitir uma interpretação mais correta do papel da mineração de dados no processo de classificação.

3.2. Geração da árvore de classificação

Primeiramente, o conjunto de dados foi inserido em um projeto do *software* eCognition (BENZ et al, 2004) para a geração dos atributos baseados em objetos que seriam usados no processo de mineração de dados. Para tanto, um processo de segmentação em duas etapas foi realizado.

Na primeira segmentação utilizou-se o algoritmo Multiresolução combinando o MDE-HAND, declividade e curvatura horizontal, com os parâmetros: Escala 100, Forma 0,5 e Compacidade 0,5. Na segunda segmentação fez-se uso do algoritmo de Diferença Espectral, com uma diferença máxima de 7 entre as alturas médias dos objetos no MDE-SRTM. Do conjunto de objetos obtidos com os dois tipos de segmentação, foram extraídos os atributos descritos na Tabela 1, que representam parâmetros estatísticos geomorfométricos, parâmetros geométricas e de textura dos objetos.

Tabela 1. Lista de atributos utilizados no processo de mineração de dados.

Atributos geométricos	Área
	Comprimento do eixo principal
	Ajuste retangular
	Ajuste elíptico
Atributos geomorfométricos	Maior valor dentro do objeto (MDE-SRTM)
	Maior valor dentro do objeto (MDE-HAND)
	Maior valor dentro do objeto (Declividade)
	Maior valor dentro do objeto (Curvatura)
	Menor valor dentro do objeto (MDE-SRTM)
	Menor valor dentro do objeto (MDE-HAND)
	Menor valor dentro do objeto (Declividade)
	Menor valor dentro do objeto (Curvatura)
	Média dos valores (MDE-SRTM)
	Média dos valores (MDE-HAND)
	Média dos valores (Declividade)
	Média dos valores (Curvatura)
	Desvio padrão dos valores (MDE-SRTM)
	Desvio padrão dos valores (MDE-HAND)
Desvio padrão dos valores (Declividade)	
Desvio padrão dos valores (Curvatura)	
Atributos texturais	GLCM Homogeneidade (em todas as direções) (MDE-SRTM)
	GLCM Homogeneidade (em todas as direções) (MDE-HAND)
	GLCM Homogeneidade (em todas as direções) (Declividade)
	GLCM Homogeneidade (em todas as direções) (Curvatura)

Para o processo da construção da árvore de classificação, foram escolhidas aleatoriamente 3000 amostras de cada classe, com base nos mapas de referência (Hess et al. 2003), para cada uma das áreas de estudo. Cada uma dessas amostras consistia em um identificador da amostra, a identificação à qual classe essa amostra pertencia no mapa de referência, e os valores de cada um dos atributos extraídos.

A tabela das amostras foi introduzida no *software* WEKA para a mineração dos dados. Foi escolhido o algoritmo de árvore de decisão J48 pela facilidade na compreensão dos resultados e na implementação da árvore de decisão criada por esse algoritmo no eCognition para a geração da classificação final. Esse algoritmo se baseia no conceito de entropia da informação para encontrar os atributos que aperfeiçoem a árvore de decisão (TAN et al, 2009).

Esse algoritmo foi rodado com os seguintes parâmetros: Fator de confiança 0,2 e Número mínimo de objetos por folha de 150 amostras. O modelo de árvore obtido foi testado pelo sistema de validação cruzada, com dez subconjuntos das amostras sendo alternadamente utilizados como base para a construção da árvore e para a sua validação. Foi criada uma árvore para cada área de estudo, seguindo essa metodologia, e a partir dessas árvores foram geradas novas classificações para a área de estudo.

3.3. Mapeamento das áreas alagáveis

Para mapear as áreas alagáveis, as árvores de decisão obtidas com a mineração de dados foram programadas no *software* eCognition, fazendo uso dos mesmos procedimentos de segmentação utilizados para a geração das amostras. Ao contrário da classificação com seleção manual de atributos, a classificação baseada na mineração de dados não foi submetida ao processo de correção para eliminar pequenos objetos. Desta forma, foram avaliados os resultados obtidos sem alteração.

3.4. Validação das classificações

A partir dos mapas de referência, foi calculado o valor do índice *Kappa* pelo método Monte Carlo da classificação baseada na mineração de dados de cada área de estudo. Nessa análise, foi definido como tamanho da amostra 150 pixels para cada classe, 450 no total, e 5.000 repetições do cálculo do índice *Kappa*. Desse conjunto de 5.000 valores, é determinado um intervalo de credibilidade de 95%, isto é, são descartados os 2,5% maiores e os 2,5% menores valores obtidos. Quanto maior o valor desse intervalo, maior a correspondência esperada entre o mapa gerado e o de referência. Quanto menor for a amplitude desse intervalo, menor é a variância dos dados e, portanto, mais consistente é o resultado. Essa consistência é indicativa de que a classificação tem desempenho semelhante em toda a área classificada.

4. Resultados e discussão

As árvores de decisão para as áreas teste são mostradas nas Figuras 4 e 5. Na Figura 6 é apresentado o resultado da classificação baseada nessas árvores de decisão.

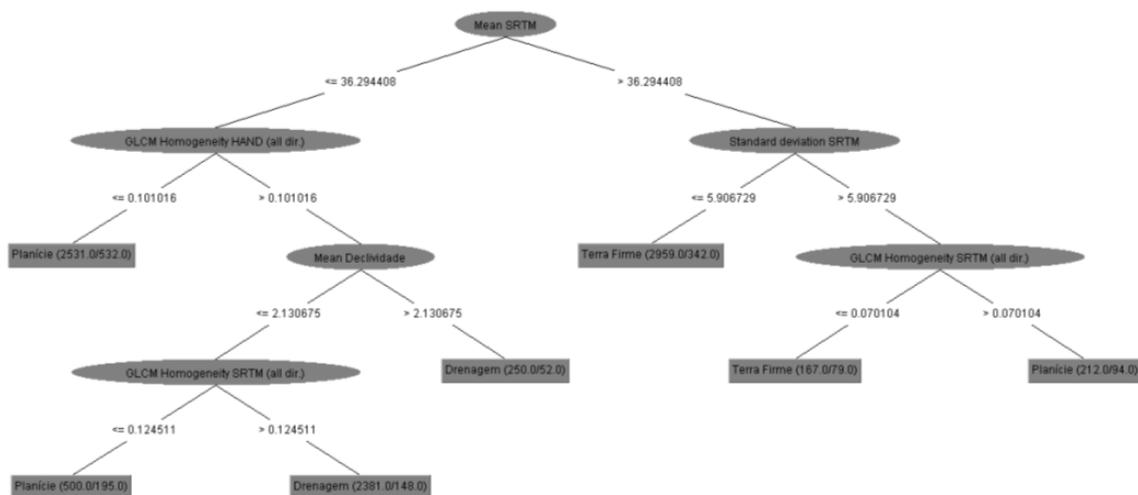


Figura 4. Árvore de decisão construída pelo processo de mineração de dados para a área teste de Codajás.

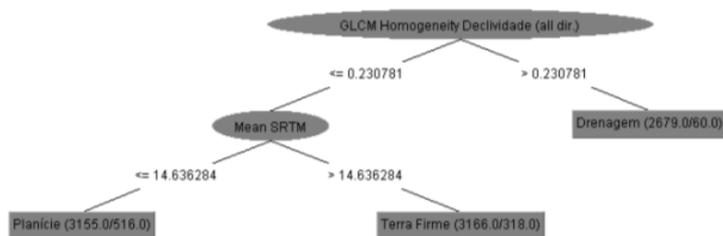


Figura 5. Árvore de decisão construída pelo processo de mineração de dados para a área teste de Óbidos.

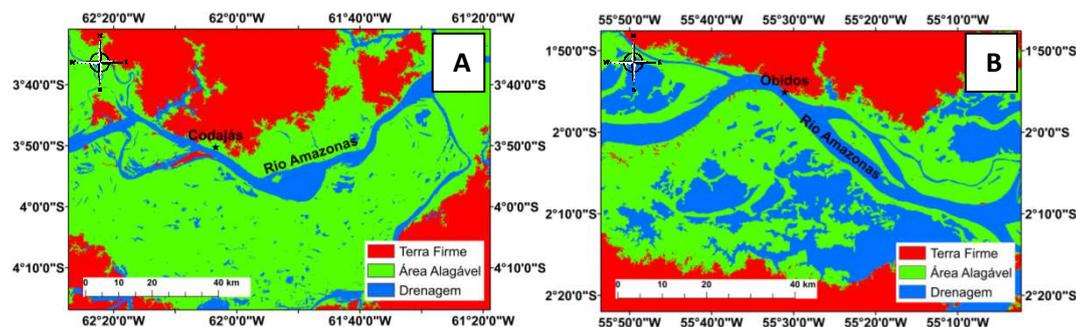


Figura 6. Mapa das áreas alagáveis de Codajás (A) e Óbidos (B), elaborado com mineração de dados e GEOBIA.

Como pode ser observado pelas figuras 4 e 5, as árvores de decisão são bastante distintas, apesar de terem sido criadas com o mesmos algoritmos, configurações e atributos. A árvore de Óbidos é semelhante àquela construída apenas com os atributos GLCM Homogeneidade (Declividade) e Média do SRTM, selecionados manualmente. Essa simplicidade da árvore reflete as condições das áreas alagáveis da região de Óbidos, que apresenta poucos afluentes de pequeno porte e lagos de pequena dimensão, que são os mais difíceis de serem classificados corretamente. Além disso, o curso do rio e a planície de inundação são mais retilíneos, ou seja, com menos variação do curso devido a processos erosivos e deposicionais.

Também contribuiu para a maior correspondência entre o mapa gerado e o de referência, o desnível altimétrico mais acentuado no contato entre a planície de

Inundação e o terraço (Terra Firme). Esse desnível é superior a 10 m na maior parte do limite entre Planície de Inundação e Terra Firme, e frequentemente superior a 20 m.

Já a árvore de decisão para Codajás é mais complexa como mostra a Figura 4, como seria de se esperar, pelas características desta área. Na região de Codajás, a planície de inundação é menos confinada do que em Óbidos, e a presença de muitos canais abandonados e pequenos lagos são fatores que dificultam a classificação, dificuldades que o algoritmo de classificação tenta superar com o aumento no número de folhas da árvore de decisão. Além disso, essa área, por apresentar taxas de erosão e deposição de sedimentos superiores em relação à Óbidos, tem uma transição mais gradual entre a planície de inundação e as áreas não inundáveis, especialmente na margem esquerda.

A Tabela 2 mostra os resultados da análise de Monte Carlo para a classificação baseada na seleção manual de atributos (ALFAYA, 2012) e naquela que utilizou o método automático por mineração de dados. Os resultados para a região de Óbidos, a partir de ambos os métodos de classificação são muito semelhantes, devido às características de sua planície de inundação, destacadas anteriormente.

Tabela 2. Resultado da análise de Monte Carlo obtido da classificação baseada na seleção manual de atributos (ALFAYA, 2012) e da aplicada com atributos selecionados automaticamente, por mineração de dados (WEKA).

Área Teste	Resultado da classificação com seleção manual de atributos			Resultado da classificação com seleção automática de atributos		
	Intervalo de credibilidade (95%)	kappa mínimo	kappa máximo	Intervalo de credibilidade (95%)	kappa mínimo	kappa máximo
Codajás	0,693 - 0,797	0.647	0.847	0,733 - 0,830	0.693	0.877
Óbidos	0,810 - 0,890	0.783	0.923	0,810 - 0,893	0.733	0.917

Para Codajás, no entanto, o resultado da classificação por mineração de dados apresenta um intervalo de credibilidade notadamente mais alto do que aquele encontrado com a classificação que utilizou atributos selecionados manualmente. Porém, em termos de amplitude do intervalo, os resultados são semelhantes. Isso indica que a árvore de decisão construída com a mineração de dados é mais adequada na classificação de áreas alagadas de planície de inundação mais complexas como é a de Codajás.

5. Conclusões

A utilização de métodos de classificação por mineração de dados pode ser uma alternativa interessante no mapeamento de áreas alagáveis, especialmente aquelas de maior complexidade morfológica, como a da área de estudo Codajás, para as quais a seleção manual dos atributos é uma tarefa difícil e demorada. Para planícies de inundação com as características de Óbidos, os resultados obtidos mostraram que não existe diferença significativa entre o método manual e automático de seleção de atributos, embora este seja menos trabalhoso. Com a mineração de dados é possível utilizar os mesmos dados e algoritmos na construção de árvores de decisão específicas para cada tipo de área de estudo. Com isto, é possível obter melhores resultados.

6. Referências

- Alfaya, F. A. V. S. **Mapeamento de áreas alagáveis da calha Solimões/Amazonas utilizando análise de imagens baseada em objeto com dados MDE-SRTM**. 2012. 77 p. (sid.inpe.br/mtc-m19/2012/04.10.11.22-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.
- Barbosa, C. C. F. **Sensoriamento remoto da dinâmica de circulação da água do sistema planície de Curuai/Rio Amazonas**. 2005. 285 p. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2005.
- Benz, U. C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. **ISPRS Journal of Photogrammetry & Remote Sensing**, n. 58, p. 239-258, 2004.
- Hess, L. L.; Melack, J. M.; Novo, E. M. L. M.; Barbosa, C. C. F.; Gastil, M. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. **Remote Sensing of Environment**, v. 87, p. 404–428, 2003.
- Jain. A. K., Robert P.W. Moa D., Moa. J. Statistical Pattern Recognition: A Review. **IEEE Trans. On Pattern Analysis And Machine Intelligence**, v. 22, n.1, 2000.
- Junk, W. Os recursos hídricos da Amazônia. In PAVAN, C., & ARAÚJO M. C. (Eds.), **Uma Estratégia Latino-americana Para a Amazônia Brasília**, Brasil: Ministério do Meio Ambiente, Recursos Hídricos e Amazônia Legal, p. 247–259, 1996.
- Mertes L. A. K.; Dunne T.; Martinelli L. A. Channel floodplain geomorphology along the Solimões-Amazon river, Brazil. **Geological Society of America Bulletin**, v.108, n.9, p.1089-1107, 1996.
- Oliveira, J. A.; Dutra, L. V.; Rennó, C. D. Aplicação de métodos de extração e seleção de atributos para classificação de regiões. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 12., 2005, Goiânia. **Anais...** São José dos Campos: INPE, 2005. Artigos, p. 4201-4208. CD-ROM, On-line. ISBN 85-17-00018-8. Disponível em: <<http://marte.dpi.inpe.br/col/ltid.inpe.br/sbsr/2004/11.20.21.24/doc/4201.pdf>>. Acesso em: 05 nov. 2012.
- Rabus, B.; Eineder M.; Roth A.; Bamler R. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. **Photogrammetry and Remote Sensing**. v. 57, p. 241-262, 2003.
- Rennó, C.D.; Nobre, A.D.; Cuartas, L.A.; Soares, J.V.; Hodnett, M.G.; Tomasella, J.; Waterloo, M.J. Hand, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazônia. **Remote Sensing of Environment**, v.112, p. 3469-3481, 2008.
- Renó, V.F.; Novo, E. M. L. M.; Suemitsu, C.; Rennó, C.D.; Silva, T. S. F. Assessment of deforestation in the Lower Amazon floodplain using historical Landsat MSS/TM imagery, **Remote Sensing of Environment**, v. 115, p. 3446-3456, 2011.
- Shimabukuro, Y. E.; Novo, E. M. L. M.; Mertes, L. K. Amazon River mainstem floodplain Landsat TM digital mosaic. **International Journal of Remote Sensing**, v.23, n.1, p. 57-69, 2002.
- Tan, P.; Steinbach, M.; Kumar, V. **Introdução ao Data Mining: Mineração de Dados**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009.
- Valeriano, M. M.; Rossetti, D. F. Topodata: Brazilian full coverage refinement of SRTM data. **Applied Geography**. v. 32, p. 300-309, 2012.
- Witten, I.H., Frank, E., Trigg, L., Hall, M., Holmes, G. & Cunningham, S.J.. **Weka: Practical machine learning tools and techniques with Java implementations**. (Working paper 99/11). Hamilton, New Zealand: University of Waikato, Department of Computer Science.1999.