

## **Avaliação do mapeamento de manguezais na Região Metropolitana do Rio de Janeiro através da integração entre GEOBIA e Mineração de Dados**

Marcus Vinícius Alves de Carvalho<sup>1,2</sup>  
Gabriella Ferreira da Silva<sup>2</sup>  
Carla Bernadete Madureira Cruz<sup>2</sup>  
Paula Maria Moura de Almeida<sup>2,3</sup>

<sup>1</sup> Universidade Federal Fluminense - UFF  
Laboratório de Geografia Física - LAGEF / Instituto de Geociências - IGEO  
Av. General Milton Tavares de Souza, s/nº - Campus da Praia Vermelha  
Gragoatá - Niterói, RJ - Brasil  
CEP: 24.210-346  
marcus\_carvalho@id.uff.br

<sup>2</sup> Universidade Federal do Rio de Janeiro - UFRJ  
Laboratório ESPAÇO - Sensoriamento Remoto e Estudos Ambientais / Instituto de Geociências - IGEO  
Av. Brigadeiro Trompowski, s/nº - CCMN  
Ilha do Fundão - Rio de Janeiro, RJ - Brasil  
CEP: 21.949-900  
{gabriella.geoufrj, carlamad}@gmail.com

<sup>3</sup> Universidade do Estado do Rio de Janeiro - UERJ  
NEMA - Núcleo de Estudos em Manguezais / Faculdade de Oceanografia  
Rua São Francisco Xavier, 524 - Campus Negrão de Lima  
Maracanã - Rio de Janeiro, RJ - Brasil  
CEP: 20.550-013  
almeida.pmm@gmail.com

**Abstract.** The objective of this paper is to evaluate the integration between GEOBIA and Data Mining applied to the mapping of mangroves in the Metropolitan Region of Rio de Janeiro (Brazil). Seven experiments with different spectral attributes were generated from TM / Landsat-5 images (two different seasonal periods). The best result was the sixth experiment (Linear spectral mixture model + NDVI) with Kappa Index: 0.77 and 0.75.

**Palavras-chave:** remote sensing, digital image processing, data mining, GEOBIA, sensoriamento remoto, processamento digital de imagens, mineração de dados, GEOBIA.

### **1. Introdução**

Devido ao caráter sinóptico, holístico, multitemporal, multiespectral e ainda pelo acesso gratuito, as imagens de baixa e média resolução espacial provenientes de Sensoriamento Remoto Orbital caracterizam-se como importantes instrumentos para o monitoramento, gestão e planejamento ambiental.

Assim, os mapeamentos do ecossistema manguezal são extremamente relevantes porque ele desempenha o papel de exportador de matéria orgânica para o estuário, suas raízes funcionam como filtros na retenção dos sedimentos e ainda porque é no manguezal que algumas espécies aquáticas e terrestres se reproduzem e se refugiam. Possuem ainda valor ecológico e econômico.

No entanto, para que sejam extraídas informações de forma mais eficiente e consistente a partir das imagens orbitais sobre o meio ambiente são necessárias técnicas de processamento digital de imagens de sensores remotos (PDI).

Nesta perspectiva, a GEOBIA integrada à Mineração de Dados tende a auxiliar consideravelmente o monitoramento ambiental, uma vez que na primeira os classificadores buscam aprimorar o desempenho dos processos automatizados de classificação de imagens e na segunda são aplicados algoritmos específicos que atendem ao desenvolvimento de métodos e técnicas para “compreender” e extrair padrões em grandes quantidades de dados. Os recursos disponíveis nesses algoritmos envolvem técnicas de estatística, recuperação de informação, inteligência artificial e reconhecimento de padrões.

Assim, o presente trabalho tem por objetivo avaliar a integração entre GEOBIA e Mineração de Dados no mapeamento temático dos manguezais presentes na Região Metropolitana do Rio de Janeiro através de imagens TM/Landsat 5 de épocas sazonais distintas (período seco e período úmido).

## **2. Fundamentação teórica**

### **2.1. Manguezal**

O manguezal é conceituado como um ecossistema costeiro, situado nas regiões tropicais e subtropicais, ocorrendo junto a desembocadura de rios, estuários e lagunas costeiras, até onde houver influência de marés. Os manguezais são ecossistemas de transição entre os ambientes oceânico e terrestre (SCHAEFFER-NOVELLI, 1989).

A importância desse ecossistema altamente produtivos, devido à grande quantidade de matéria orgânica gerada neste ambiente e liberada para as águas costeiras na forma de detritos, compoem a base alimentar de várias espécies de caranguejos, camarões e peixes de valor comercial (ODUM, 1971).

Além da exportação de detritos para o estuário, Maciel (1989) cita a importância do manguezal como fixador de sedimentos e habitat para as espécies de valor comercial. Diegues (1990) destaca outras importantes funções desempenhadas pelo manguezal, tais como: proteção da linha de costa, berçário para inúmeras espécies de peixes, crustáceos e moluscos, além da manutenção da qualidade da água.

### **2.2. GEOBIA - Análise de Imagem Baseada em Objeto Geográfico**

Os classificadores que fazem uso da abordagem conhecida como Análise de Imagem Baseada em Objeto Geográfico buscam aprimorar o desempenho dos processos automatizados de classificação de imagens de Sensores Remotos. Eles possuem recursos que permitem sistematizar e reproduzir o conhecimento do intérprete humano. BLASCHKE et al., (2000) destacam que é possível que a semântica seja desenvolvida apoiada em parâmetros físicos e de conhecimento sobre relacionamentos espaciais dos objetos.

Segundo Hay e Castilla (2008), GEOBIA é uma subdisciplina da Ciência da Informação Geográfica (*GIScience*) dedicada ao desenvolvimento de métodos automatizados para a significativa divisão de imagens de sensoriamento remoto em objetos-imagem, e para avaliar suas características por meio de escalas espaciais, espectrais e temporais, com o intuito de gerar novas informações geográficas em SIG. DEFINIENS (2007) destaca que a utilização do conceito de objeto é peça-chave neste tipo de análise de imagens, pois se parte do princípio de que a informação semântica necessária para a interpretação de uma imagem não está presente apenas no pixel, e sim em objetos da imagem e nas relações existentes entre eles.

### **2.3. Mineração de dados**

Mineração de dados é o nome dado ao conjunto de técnicas que permite a extração de conhecimentos a partir de grandes volumes de dados. O processo de transformação de dados em informações e conhecimento é conhecido como descoberta de conhecimento em base de dados (*KDD - Knowledge Discovery in Databases*). A mineração de dados trata da conversão de dados ou informações (que podem ter sido pré-processadas) em padrões (SANTOS, 2012).

Segundo Goldschmidt e Passos (2005), 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 interessantes 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 (como, por exemplo, a forma, extensão) e as relações espaço-temporais entre os objetos e eventos, além dos atributos não-espaciais de interesse na mineração de dados tradicional (MILLER; HAN, 2009).

### 3 Metodologia

A metodologia adotada neste trabalho é apresentada na Figura 1.

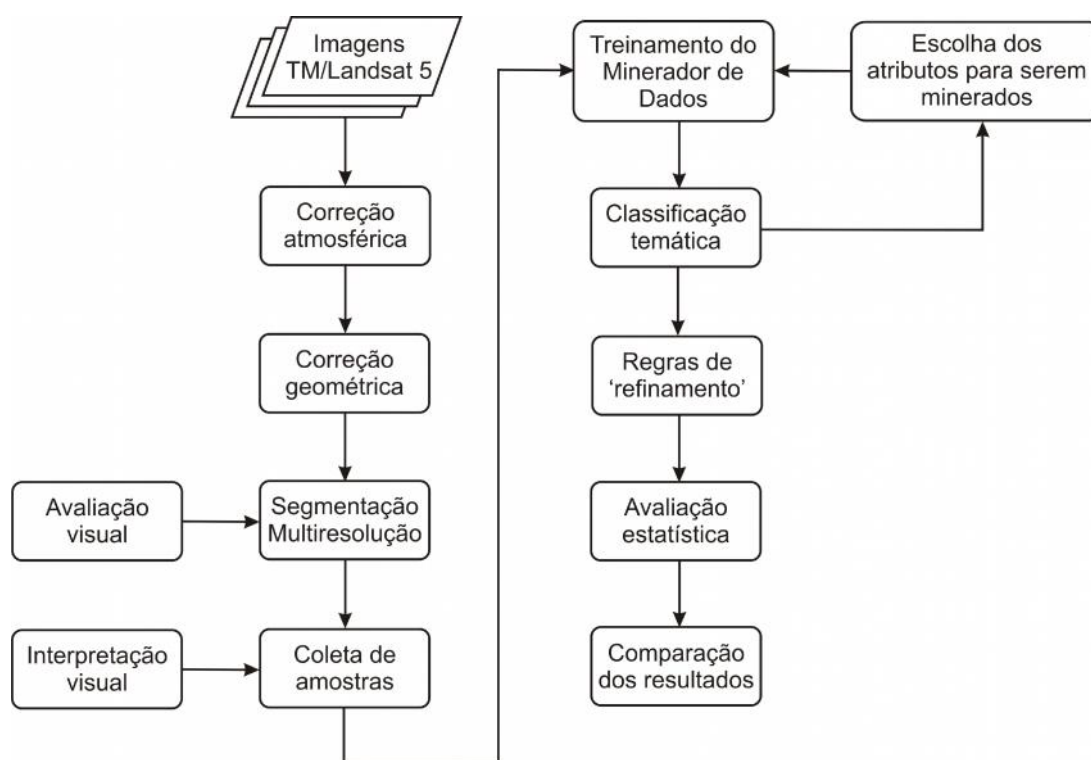


Figura 1 - Fluxograma metodológico.

A área de estudo, Região Metropolitana do Rio de Janeiro, é compreendida em sua maior parte pela cena 217/076 do sensor *Thematic Mapper* (TM) do satélite Landsat-5.

As imagens TM/Landsat 5 foram adquiridas gratuitamente no Centro de Dados de Sensoriamento Remoto (CDSR) da Divisão de Geração de Imagens (DGI) do Instituto Nacional de Pesquisas Espaciais (INPE). As imagens são datadas de 23/02/2007 (período úmido) e 02/08/2007 (período seco). Adotou-se neste estudo imagens do ano de 2007 porque o Laboratório ESPAÇO/UFRJ através do Projeto Prioridade Rio já dispõe de um mapeamento prévio (verdade de campo).

A primeira etapa consistiu na correção geométrica das imagens. Para tal fez-se uso do módulo *Registration* (opção *Image to Image*) da plataforma ENVI. Assim, tomou-se como base uma imagem OLI/Landsat 8 no site do Serviço Geológico dos Estados Unidos (USGS)

para corrigir espacialmente a imagem TM/Landsat-5. Vale destacar que as imagens OLI e TIRS já têm sido disponibilizadas ortorretificadas pelo USGS.

A etapa seguinte consistiu na correção atmosférica, uma vez que, a atmosfera interage com a radiação eletromagnética, provocando significativas alterações no fluxo radiante proveniente do alvo. Para tal utilizou-se o aplicativo ATCOR presente no software *PCI Geomatics*.

Em seguida as imagens foram segmentadas em ambiente eCognition Developer. Os parâmetros adotados foram: fator de escala: 10 / forma: 0,4 / compacidade: 0,5. Chegou-se a estes valores de forma empírica. O resultado da segmentação foi apreciado através de avaliação visual.

A etapa que segue consistiu na seleção de amostras para treinar o algoritmo de Mineração de Dados. Neste trabalho foi utilizado o CART (*Classification And Regression Tree*) que faz uso da técnica de indução por árvore de decisão. Definiu-se para este trabalho as seguintes classes: “Mangue”, “Não Mangue” e “Água”. Foram utilizadas 35 amostras de “Mangue” e 171 amostras de “Não-Mangue”.

Foram avaliados diferentes experimentos com a inserção ou retirada de atributos espectrais. Os experimentos e seus respectivos atributos são apresentados na Tabela 1.

Tabela 1 – Descrição dos experimentos e atributos utilizados.

Experimento	Atributos
01 - período seco / período úmido	Bandas espectrais (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
02 - período seco / período úmido	NDWI (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
03 - período seco / período úmido	NDVI (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
04 - período seco / período úmido	Modelo de Mistura Espectral (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
05 - período seco / período úmido	Modelo de Mistura + NDWI (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
06 - período seco / período úmido	Modelo de Mistura + NDVI (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel);
07 - período seco / período úmido	NDVI + NDWI (média, moda, quartil, desvio padrão, razão, mínimo valor do pixel e maior valor do pixel).

Após o treinamento do algoritmo e geração do modelo, procedeu-se à classificação propriamente dita. Duas regras foram aplicadas ao resultado de cada classificação com o intuito de “refinar” os resultados. Foram elas: Média do NDWI  $\geq 0,4$  e Média da Declividade  $\geq 5$ . Este último atributo foi extraído a partir do arquivo de declividade do Projeto TOPODATA (VALERIANO, 2008).

Avaliação das classificações foi realizada por meio do cálculo do Índice *Kappa* (COLGATON & CREEN, 1999).

#### 4. Resultados e discussão

Os mapas temáticos provenientes dos experimentos propostos neste trabalho (tanto período seco quanto úmido) são apresentados na Figura 2.

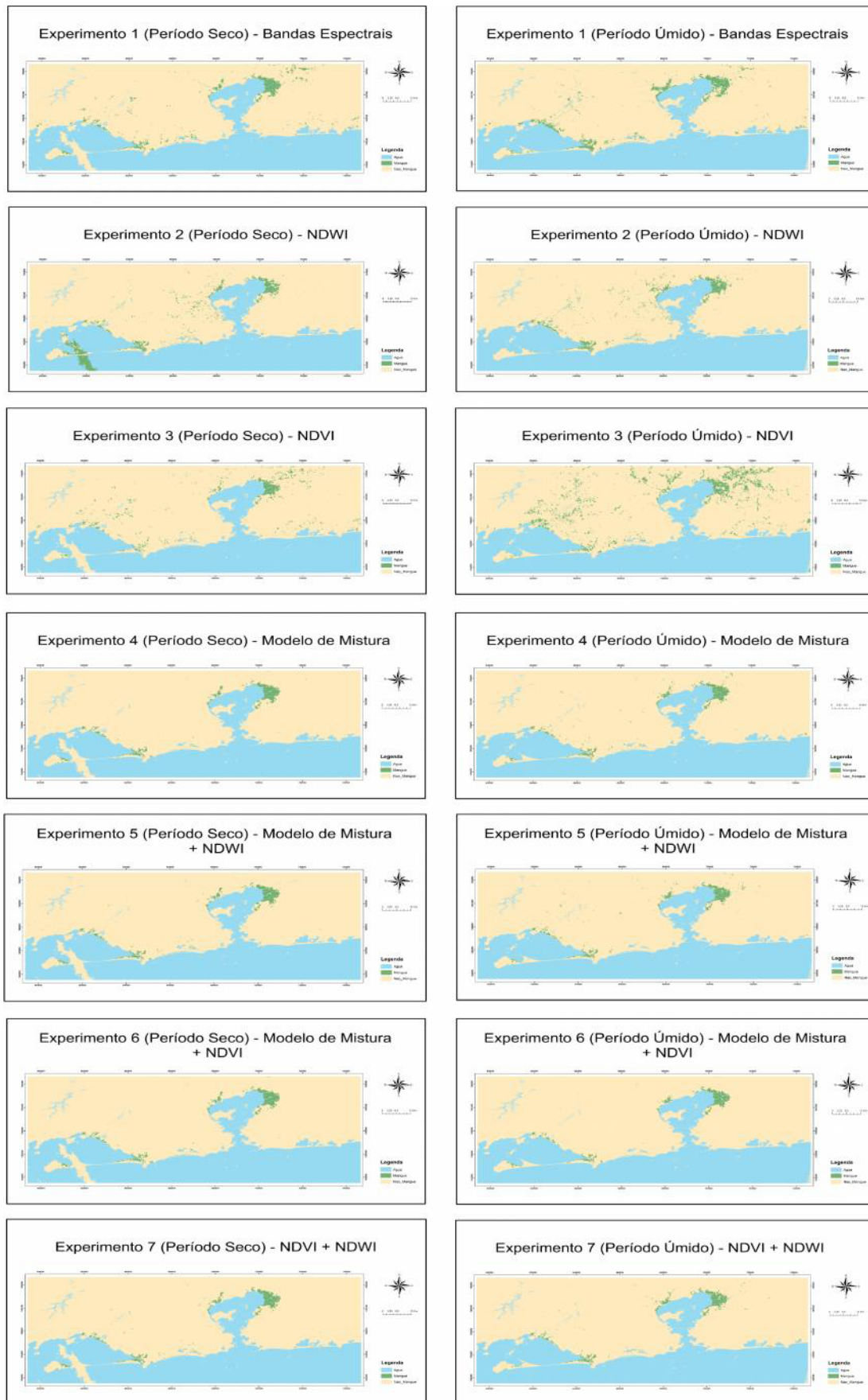


Figura 2 - Resultados dos sete experimentos propostos.

A Tabela 2 apresenta os resultados da análise da precisão para cada experimento através do Índice *Kappa*.

Tabela 2 - Precisão alcançada por cada experimento.

Experimento	Índice <i>Kappa</i>	
	Período seco	Período úmido
01	0,64	0,58
02	0,64	0,61
03	0,61	0,55
04	0,70	0,69
05	0,67	0,63
06	0,77	0,75
07	0,73	0,71

Ao analisarmos os resultados obtidos constatamos que o Experimento 1 (compostos pelos atributos provenientes de seis bandas espectrais do sensor TM, a exceção da banda do Termal) apresentou a mais baixa exatidão em ambos os períodos (seco e úmido).

No Experimento 2 avaliou-se a adoção do Índice de Diferença Normalizada da Água (NDWI). Este índice caracteriza-se por ser sensível a mudanças no teor de água da cobertura vegetal, porque a reflectância em 857 nm e 1241 nm tem propriedades de absorção de água em estado líquido semelhante, mas ligeiramente diferentes. A dispersão da luz por dosséis aumenta a fraca absorção de água líquida em 1241 nm (ITT, 2009).

O NDWI é gerado pela razão entre a subtração do verde pelo infravermelho próximo com a soma do verde pelo infravermelho próximo (Equação 1). Ele varia de -1 a 1. Assim, os valores próximos de 1 indicam relevante presença de água, os valores próximos a 0 indicam moderada presença de água e os valores próximos de -1 indicam baixa presença de água, ou seca.

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \quad (1)$$

Constatou-se superestimação e subestimação da classe “Mangue” em ambos os períodos, principalmente no período úmido. No período seco uma parte da nuvem que cobre a área chegou a ser classificada como “Mangue”.

No Experimento 3 avaliou-se a adoção do Índice Normalizado da Vegetação (NDVI) (ROUSE et al., 1973). Sua geração se dá pela diferença entre a reflectância do infravermelho próximo e a reflectância do vermelho, dividida, respectivamente, pela soma das duas reflectâncias (Equação 2). O índice gerado varia de -1 a 1. Quanto maior o valor do índice maior a presença de vegetação.

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (2)$$

Em ambos os casos ocorreu a superestimação e subestimação da classe “Mangue” (outros tipos de vegetação foram erroneamente classificados), porém entre todos os experimento o período úmido foi o que apresentou a mais baixa exatidão.



O Experimento 4 consistiu na adoção do Modelo Linear de Mistura Espectral (MLME). Segundo INPE (2014), o MLME baseia-se na premissa que quando um sensor remoto observa a cena, a radiância detectada é a integração (mistura) de todos os objetos, denominados componentes da mistura, contidos no elemento de cena. Deste modo, por meio da extração das respostas individuais de cada alvo a partir de amostras de pixels puros desses alvos, também chamados de *endmembers*, pode-se modelar o peso com que cada alvo está contribuindo para o sinal de cada pixel e, assim realizar uma composição deste sinal em imagens-fração (ROBERTS et al., 1998).

Este experimento obteve uma boa exatidão. Para ambos os casos, o algoritmo de mineração de dados selecionou o atributo de Razão da imagem fração Vegetação (ramo inicial) para definir o que não é mangue. Vale destacar que as imagens-fração foram geradas no aplicativo SPRING (CÂMARA et al., 1996).

O Experimento 5 consistiu na mineração dos atributos provenientes do Modelo de Mistura mais os atributos do NDWI. O período seco alcançou um melhor resultado que o período úmido. Já no período úmido ocorreu uma considerável classificação de objetos como “Mangue” quando na verdade não o são.

No Experimento 6 (Modelo de Mistura Espectral mais NDVI), os resultados foram os mais satisfatórios de todos os experimentos. Vale destacar que neste experimento ocorreu uma menor superestimação e subestimação da classe “Mangue” no período úmido, assim como ocorreu no Experimento 5.

O Experimento 7 consistiu na integração dos atributos do NDVI mais os atributos do NDWI. Este experimento apresentou bons resultados.

Etapas importantes para ter alcançado estes valores foi a proposição de regras para realizar o refinamento das classificações (estratégia possível em GEOBIA). A Figura 3 demonstra o impacto deste recurso neste trabalho.

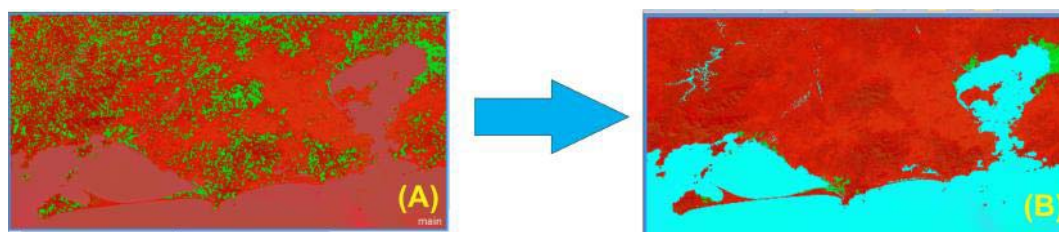


Figura 3 – (A) Comparação entre o resultado do modelo de classificação sem a aplicação das regras de refinamento (B) e o resultado do mesmo modelo de classificação já tendo sido aplicadas as regras de refinamento.

## 5. Conclusão

Mediante os resultados apresentados, podemos concluir que o objetivo do presente trabalho foi atendido, o de avaliar as imagens TM/Landsat 5 por meio da integração entre GEOBIA e Mineração de Dados no mapeamento dos manguezais. O Experimento que obteve o melhor resultado entre os sete propostos foi o referente ao Modelo de Mistura + NDVI.

Este trabalho não busca esgotar o assunto, contudo, novos testes com mais atributos deverão ser realizados. A disponibilização de imagens OLI e TIRS/Landsat 8 abrem mais perspectivas. Vale destacar que até o presente momento, o SPRING não dispõe de modelos para gerar o MLME, justificando assim a não adoção deste novo dado.

## 6. Agradecimentos

Os autores expressam seus agradecimentos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa de estudos ao primeiro autor e ao

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de estudos a segunda e quarta autora respectivamente.

## 7. Referências bibliográficas

- BLASCHKE, T.; LANG, L.; LORUP, E.; STROBL, J.; ZEIL, P. Object-oriented image processing in an integrated GIS/Remote sensing environment and perspectives for environmental applications. In: CREMERS, A.; GERVE, K. (eds.). Environmental information for planning, politics and the public. Metropolis: Verlag, Marburg, 2000. v. 2, p. 555-570.
- CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U. M.; GARRIDO, J. C. P. Spring: Integrating Remote Sensing and GIS with Object-Oriented Data Modelling. Computers and Graphics, v.15, n.6, p.13-22, 1996.
- CONGALTON, R.; GREEN, K. Assessing the Accuracy of remotely sensed data: principles and practices. Boca Raton, FL.: CRC/Lewis Press, 1999. 137 p.
- DEFINIENS. DEFINIENS Professional 7: Reference Book. Munich: DEFINIENS (The Imaging Intelligence Company), 2007. 122 p.
- DIEGUES, A.C. Comunidades litorâneas e os manguezais do Brasil. In: Simpósio sobre ecossistemas da costa sul e sudeste brasileira, 2. Águas de Lindóia, 1990. Síntese dos conhecimentos. São Paulo, Academia de Ciências do Estado de São Paulo, publicação ACIESP 71, 1990. v.3, p.124-146.
- GOLDSCHMIDT, R.; PASSOS, E. Data mining - um guia prático. Rio de Janeiro: Elsevier, 2005. 257 p. ISBN (85-352-1877-7).
- HAY, G.J.; CASTILLA, G. Geographic Object-based Image Analysis (GEOBIA): a new name for a new discipline. In: BLASCHKE, T.; LANG, S.; HAY, G.J. (Eds) Object-based image analysis- spatial concepts for knowledge-driven remote sensing applications. Berlin: Springer-Verlag, 2008.
- INPE. Guia de usuário do SPRING, São José dos Campos, 2014.
- ITT. Guia de Usuário do ENVI, 2009.
- MACIEL, N.C. Os manguezais e as unidades de preservação no Brasil. In: Simpósio sobre ecossistemas da costa sul e sudeste brasileira, 1., Cananéia, 1987. Síntese dos conhecimentos. São Paulo, Academia de Ciências do Estado de São Paulo, Publicação ACIESP 54, 1987. v.1, p.149-172.
- MILLER, J. H.; HAN, J. Geographic data mining and knowledge discovery. 2. ed. EUA: Taylor & Francis Group, 2009.
- ODUM, W.E. 1971. Pathways of energy flow in a south Florida estuary. Sea Grant Technical Bulletin No. 7. University of Miami, Sea Grant Program (Living Resources), Miami, Florida.
- ROBERTS, S., HUSMEIER, D., REZEK, I., PENNY, W. Bayesian approaches to gaussian mixture modelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(11):1133-1142.
- ROUSE, J.W., HAAS, R.H., SCHELL, J.A. and DEERING, D.W., 1973, Monitoring vegetation systems in the Great Plains with ERTS. In 3rd ERTS Symposium, NASA SP-351 I, pp. 309–317.
- SANTOS, R. Introdução à mineração de dados com aplicações em ciências ambientais e espaciais, 2012. Disponível em: <[http://www.lac.inpe.br/ELAC/arquivos/MiniCurso\\_03ELAC2012.pdf](http://www.lac.inpe.br/ELAC/arquivos/MiniCurso_03ELAC2012.pdf)>.
- SCHAEFFER-NOVELLI, Y. Situação atual do grupo de ecossistemas: “Manguezal, Marisma e Apicum” incluindo os principais vetores de pressão e as perspectivas para sua conservação e usos sustentável. São Paulo, Brasil, 1989, p. 119.
- VALERIANO, M. M. Topodata: guia para utilização de dados geomorfológicos locais. São José dos Campos: INPE, 2008. 72 p. (INPE-15318-RPE/818). Disponível em: <<http://urlib.net/8JMKD3MGP8W/33EPEBL>>. Acesso em: 10 nov. 2014.