

Integração de dados de Sensoriamento Remoto multi resoluções para a representação da cobertura da terra em Campos Contínuos de Vegetação e classificação por árvores de decisão

Marcelo Lopes Latorre^{1,2}
João Roberto dos Santos¹
Yosio Edemir Shimabukuro¹
Dalton de Morisson Valeriano¹
Osmar Abílio de Carvalho Júnior³

1 – INPE – Instituto Nacional de Pesquisas Espaciais
Caixa Postal 515 – 12201-970–São José dos campos – SP, Brasil. Fone: (12)39456483.
E-mail:(yosio,jroberto,latav) @ltid.inpe.br

2 - CTA–Comando-Geral de Tecnologia da Aeronáutica
Praça Marechal Eduardo Gomes Nº 50, São José dos Campos, SP, Brasil.

3 - Universidade de Brasília – Departamento de Geografia – Campus Universitário Darcy
Ribeiro - ICC Norte, 70910-900, Brasília, DF, Brasil.

Abstract: This work had as general objective to develop a methodology for the dynamic's monitoring of the vegetable covering in Amazon, where a larger emphasis will be given in the integration of different sensor data and space resolutions. So much, was used for the system's development, initially as base, the technique denominated Vegetation Continuous Fields (VCF). As study area, it was chosen an area of Mato Grosso State, for presenting great representativity in relation to the Amazonian Brazilian in terms of earth covering. The methodological process consisted of establishing an exploratory analysis of the technique VCF and the development of the monitoring system, using the program ENVI and your corresponding language IDL. As result was obtained a system with capacity of obtaining data (maps), so much of 500m, as 250m of MODIS, in a period that can vary from one month to one year. Through your preliminary results and your comparison with the data obtained with the PRODES Project (year of 2002), the viability of the presented methodology was verified.

Palavras-Chave: Sensoriamento Remoto, campos contínuos, Amazônia Brasileira, sistema de monitoramento, IDL.

1. Introdução

Este trabalho tem como objetivo geral descrever o processo metodológico desenvolvido para o monitoramento da dinâmica da cobertura da terra na Amazônia, onde uma maior ênfase foi dada na integração de dados de diferentes sensores e resoluções (Terra/MODIS, Landsat5/TM, Enhanced Thematic Mapper - ETM+ e CBERS 2/CCD). Para tanto, empregou-se no seu desenvolvimento a técnica desenvolvida por Hansen et al. (2002, 2003) denominada *Vegetation Continuous Fields* (VCF) ou Campos Contínuos de Vegetação.

Nessa técnica, a floresta é descrita como porcentagem de cobertura arbórea, ao invés de floresta ou qualquer outro rótulo existente. O procedimento é automatizado e gera uma regressão (Venables & Ripley, 1994) com dados de treinamento do Satélite Landsat5/TM, ETM+ e IKONOS, coincidentes com os dados de resolução espacial ampla, onde seus produtos têm-se mostrados eficientes no monitoramento global da cobertura terrestre (Hansen et al., 2000, 2002).

Espera-se que esta metodologia auxilie e complemente, em especial, a projetos já consagrados e desenvolvidos para a Amazônia Brasileira, como é o caso do Projeto de Estimativas de Desflorestamento da Amazônia (PRODES).

1.2 Área de Estudo

O Estado do Mato Grosso (**Figura 1**) está localizado no Sul da região Amazônica (Centro-Oeste do Brasil) e apresenta uma área territorial de cerca de 900.000 km². Esta região foi escolhida por apresentar uma grande diversidade de tipos de formações vegetais, além de estar sujeita a uma alta taxa na conversão da sua cobertura vegetal, devido à utilização da agricultura mecanizada e pecuária nas áreas de cerrado, bem como o desmatamento e queimadas nas áreas de floresta.

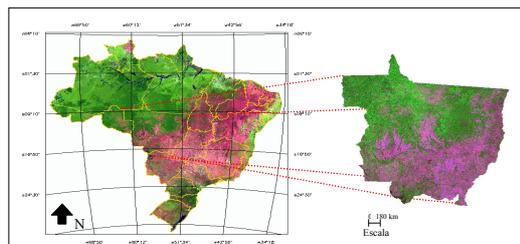


FIGURA 1 – Localização do Estado do Mato Grosso (Modificado de Anderson *et al.*, 2005, p.8).

2. MATERIAIS E MÉTODOS

2.1. Imagens e Programas Utilizados

Para a realização deste estudo foram utilizadas imagens dos seguintes sensores: (a) Landsat5/TM – banda azul (450-520nm), verde (520-600nm), vermelho (630-690nm); (b) Landsat7/ETM+ – banda azul (450-520nm), verde (530-610nm), vermelho (630-690nm); (c) CBERS 2/CCD – banda azul (450-520nm), verde (520-590nm), vermelho (630-690nm) e infravermelho próximo (770-890nm); (d) MODIS 500m – bandas: azul (459-479nm), verde (545-565nm), e infravermelho médio (1230-1250nm, 1628-1652nm e 2105-2155nm), referentes às 11 composições (32 dias) relativas ao ano de 2002; e (e) MODIS 250m – banda vermelho (620-670nm) e infravermelho próximo (841-876nm), ano de 2002. O emprego de diversos sensores teve como objetivo integrar e avaliar dados de diferentes resoluções (radiométricas, espaciais e espectrais) no desenvolvimento da metodologia proposta.

Os programas utilizados foram: (a) programa ENVI 3.6 (*Environment for Visualisation Images*) para o processamento das imagens; (b) linguagem IDL (*Interactive Data Language*) 5.6 para o desenvolvimento do programa de monitoramento e a sua interface; (c) programa *S-PLUS* empregado no treinamento das variáveis (dependentes e independentes) criadas, através de um algoritmo de regressão baseado em árvore, e aplicadas no modelo proposto; (d) programa PCI *Geomatica* 9.1 para obter os pontos para o registro das imagens de melhor resolução (TM, ETM+ e CBERS 2/CCD) com os dados MODIS (500m e 250m).

2.2. Desenvolvimento do Sistema de Monitoramento

No sistema de monitoramento para a Amazônia Legal, buscou-se desenvolver uma metodologia baseada no VCF (Defries *et al.*, 1997 e Hansen *et al.*, 2002, 2003) estabelecendo algumas modificações, como: (a) maior automatização na geração do produto em relação aos já existentes para a Amazônia Brasileira; (b) criação de um sistema aplicando as técnicas de campos contínuos, mas de maneira mais simplificada (menor número de classes), que permitisse, futuramente, gerar produtos em intervalos menores, como exemplo, quadrimestrais e mensais (dados MODIS de 500m e 250m respectivamente); (c) utilização de novos sensores como o CBERS 2/CCD; (d) facilidade na interação com o modelo

desenvolvido, uma vez que cerca de 90% de todo o processo é realizado com um único “software”, neste caso o ENVI (linguagem IDL); e (e) criação de uma interface gráfica para uma maior interação junto ao usuário.

Desta forma, o desenvolvimento do sistema de monitoramento apresentou as seguintes fases: (a) criação das variáveis independentes (as métricas - a partir dos dados MODIS 500m e 250m) e das variáveis dependentes (dados de treinamentos gerados a partir de cenas de melhor resolução espacial: ETM+, TM e CBERS 2/CCD); (b) utilização da regressão em árvore do programa *S-PLUS* e emprego na cena MODIS gerando ao final um mapa da região; e (c) criação de sua interface gráfica, já citada anteriormente.

2.3. Obtenção das Métricas e dos Dados de Treinamento

As métricas multitemporais capturam alguns pontos da variação fenológica através dos cálculos de médias (mínimas e máximas) anuais e as amplitudes das informações espectrais. Diversos tipos de métricas para a descrição da fenologia da vegetação têm sido sugeridos. Muitos autores têm obtido métricas ligadas a eventos fenológicos, baseados em uma série temporal de dados NDVI, incluindo taxa de senescência, tempo de pico de NDVI, máximo NDVI, entre outros (Reed et al., 1994).

Neste trabalho, para os dados de 500m, por exemplo, as sete bandas referentes à cobertura terrestre foram usadas como entrada: azul (459-479nm), verde (545-565 nm), vermelho (620-670 nm) e infravermelho próximo (841-876 nm) e médio (1230-1250 nm, 1628-1652 nm e 2105-2155 nm). Neste caso, dados compostos do MODIS foram transformados em métricas anuais obtidas a partir destes canais e do NDVI, as quais foram usadas como dados de entrada (variáveis independentes) para se gerar o mapa final (em porcentagem de cobertura arbórea ou em classes discretas).

A **Figura 2** exemplifica a obtenção de uma métrica para a banda do vermelho relativa ao dado MODIS de 500m (anual). Neste caso, considerando apenas as 11 composições referentes ao canal do vermelho, verifica-se que o primeiro passo é o ordenamento decrescente *pixel a pixel* de cada uma das onze composições em função do seu valor de resposta espectral. Este ordenamento irá gerar ao final onze novas composições relativas à banda do vermelho, variando do maior (1^a composição) para o menor valor (11^a composição). Uma vez estabelecido o ordenamento, dá-se início à criação da métrica, através da execução de uma média em relação às novas composições criadas.

Assim, quando se fala em métrica REDmean19, por exemplo, interpreta-se como uma média da primeira à nona composição (já ordenada) do canal vermelho. Cabe salientar que, este processo foi aplicado, também, aos demais canais.

O próximo passo está na obtenção dos futuros dados de treinamento. Os dados de treinamento são obtidos, inicialmente, por meio da classificação e interpretação de imagens de melhor resolução considerando áreas mais homogêneas (floresta e não floresta e água / áreas inundáveis ou qualquer outra classe). Cabe salientar que, neste trabalho, como a ênfase maior estava no desenvolvimento metodológico, optou-se por definir apenas três classes, ou seja: floresta, não floresta e áreas inundáveis. Para tanto, empregaram-se imagens dos sensores: CBERS 2/CCD, Landsat5/TM e Landsat7/ETM+.

A próxima etapa na obtenção dos dados de treinamento foi o registro com os dados MODIS. Todo este processo é realizado utilizando-se dos programas ENVI (programa em IDL criado especialmente para esta tarefa) e o PCI 9.1 Geomática. O programa PCI foi utilizado apenas para coletar os pontos de controle necessários ao registro das cenas. Em seguida, estes pontos são inseridos durante o registro no ENVI.

Nesta fase de registro as classificações em imagens de melhor resolução são agregadas às piores (dados de 500 m), rotulando cada estrato em função dos valores médios de cobertura

arbórea (Por exemplo: 0%, 25%, 50% e 80%); sendo em seguida, estabelecido um valor médio sobre as células de resolução mais ampla (**Figura 3**). Nesta etapa é estabelecida uma mesma dimensão da grade para as métricas e dados de treinamento.

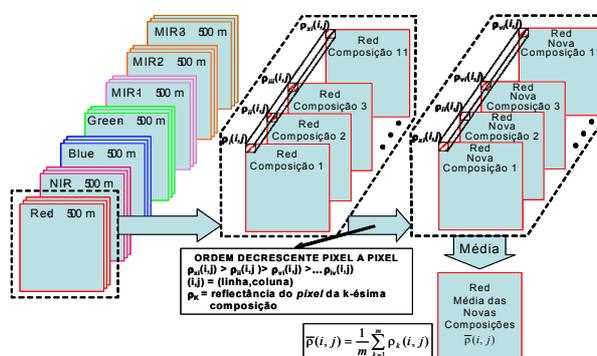


FIGURA 2 – Exemplo na obtenção de métricas para o canal vermelho do dado MODIS 500m (anual).

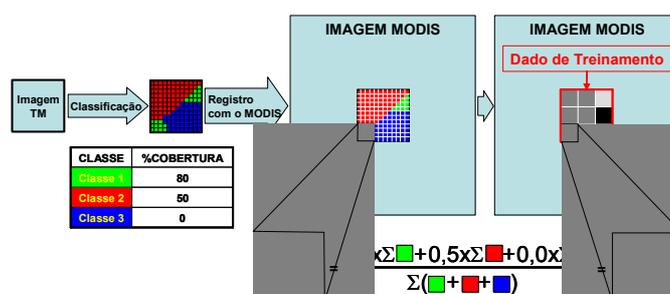


FIGURA 3 – Exemplo de passos empregados na obtenção de um dado de treinamento.

As amostras seleccionadas e suas informações são convertidas em arquivos .TXT, que são inseridos no programa *S-PLUS* para o treinamento da árvore de decisão.

2.4. Árvore de Decisão

Uma árvore de decisão tem a função de dividir um conjunto de treinamento, até que cada subconjunto obtido deste particionamento contenha casos de uma única classe. Para atingir esta meta, esta técnica examina e compara a distribuição de classes durante a construção da árvore. Os resultados obtidos são dados organizados de maneira compacta, que são utilizados para classificar novos casos (Friedl & Brodley, 1997).

Cabe salientar que, empregou-se o *S-PLUS*, em função de algumas de suas características: flexibilidade, simplicidade intuitiva e eficiência computacional, e também, por possuir uma elevada aceitação nos problemas envolvendo a classificação de cobertura terrestre, em especial para escalas regional e global (Strahler *et al.*, 1999).

Resumidamente, este algoritmo é implementado da seguinte forma. Duas amostras de pixels de treinamento são definidas a partir de um conjunto de dados de treinamento. Este processo denominado de resubstituição estabelece que uma amostra seja usada para desenvolver (GTR) a árvore e a outra amostra para “podá-la” (PTR). Esta “poda” se faz necessária, porque este tipo de algoritmo é muito robusto, possuindo a capacidade de delinear um simples *pixel* em um espaço espectral.

Através de um conjunto à parte de dados de treinamento, uma árvore mais generalizada pode ser criada. Esta generalização é obtida através de uma segunda amostra, logo abaixo da árvore inicial. Com os dados em cascata, logo abaixo da árvore, a soma geral dos quadrados

começa a aumentar. Isto indica uma superestimação de uma árvore inicial. Neste método, a “poda” é executada, não quando as somas dos quadrados começavam a aumentar, mas sim, onde os nós adicionais representavam uma redução da soma geral dos quadrados para estes dados. O resultado final é uma hierarquia interpretada de divisões, as quais, quando seguidas, permitem uma distinção do produto gerado (**Figura 4**).

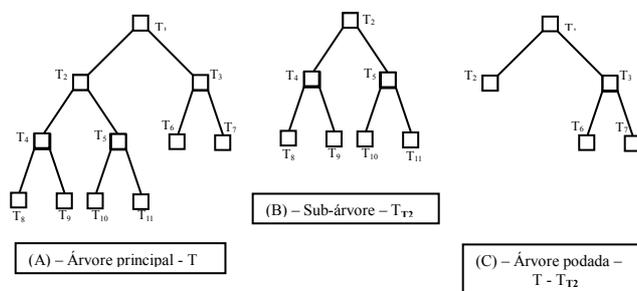


FIGURA 4 – Exemplo de poda. (A) árvore completa; (B) sub-árvore; e (C) árvore final após a poda. (Modificado de Breiman 1984, p. 64).

2.5. Obtenção dos Mapas Finais

Nesta fase, os resultados obtidos pela regressão são convertidos em linguagem IDL e aplicados à cena do MODIS, gerando-se ao final, um mapa da área de estudo em classes com valores discretos ou em porcentagem de cobertura arbórea.

Cabe salientar que, os melhores resultados foram separados empregando-se uma interpretação visual, complementada por uma análise comparativa (matriz de confusão – acurácia global e coeficiente kappa) utilizando-se regiões de interesse (ROIS), geradas no programa ENVI, a partir de uma análise das métricas obtidas durante o processo. Neste caso, separaram-se para toda a cena MODIS as regiões correspondentes às classes de floresta, não floresta e água / áreas inundáveis.

3. RESULTADOS E DISCUSSÕES

As **Figuras 5** abordam um exemplo de resultado obtido com o modelo de monitoramento descrito (dados 500m – valores discretos e/ou porcentagem de cobertura); e a **Figura 6** um resultado obtido para os dados de 250m do MODIS, o qual procurou-se enfatizar o grau de detalhamento obtido com o modelo, apesar do número reduzido de classes. A obtenção de produtos com os dados 250m, embora não fosse o foco principal do trabalho, teve como objetivo geral verificar, a viabilidade futura do seu emprego no monitoramento da cobertura terrestre, com o modelo proposto, uma vez que se conseguiu gerar produtos mensais e bimestrais da região.

O final desta etapa consistiu em separar os melhores resultados obtidos anteriormente, para se realizar uma análise comparativa com os produtos obtidos pelo Projeto PRODES. Esta consistiu em estabelecer algumas adaptações, uma vez que, os produtos apresentavam projeções, resoluções e metodologias diferenciadas. Para tanto: (a) os mapas foram colocados na mesma projeção, utilizando como referência a definida para os dados MODIS; (b) Definiu-se das mesmas resoluções, uma vez que os dados PRODES estavam na resolução de 120m e estes dados MODIS (empregados como base) em 467.563m; (c) estabelecimento das mesmas classes; e (d) as cenas foram recortadas de maneira correspondente, ou seja, na mesma grade ou área de atuação.

Considerando a não edição do produto obtido pelo modelo, juntamente, com as inúmeras adaptações necessárias para se conseguir estabelecer esta comparação, observa-se na **Figura**

7, que o modelo apresentado mostrou-se viável, tendo em vista os resultados, bem como os valores de acurácia global apresentados.

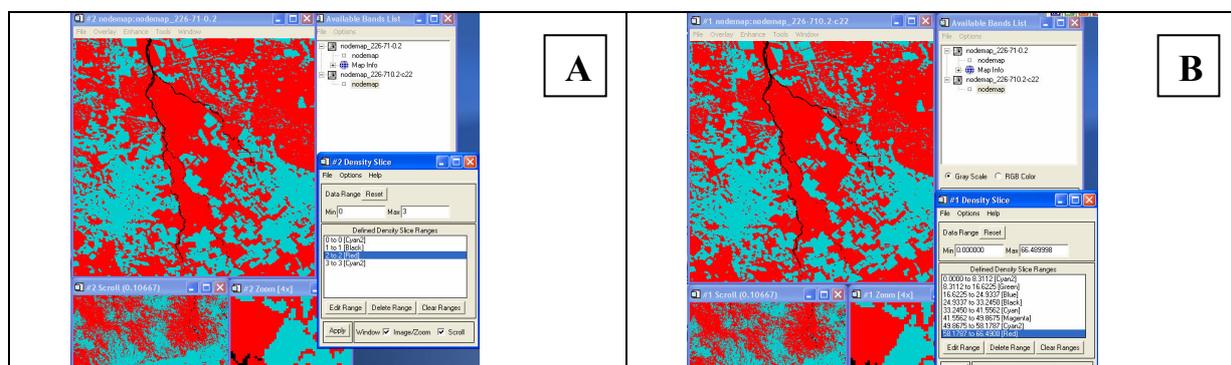


FIGURA 5 – A) Exemplo de mapa final da região do Mato grosso, utilizando-se dados MODIS 500m para o ano de 2002, contendo três classes: vermelho – floresta; ciano - não floresta e preto – água / áreas inundáveis. B) Mesma região contendo três classes em valores de porcentagem arbórea, definida nos dados de treinamento empregados: vermelho – floresta; ciano - não floresta e preto – água / áreas inundáveis.

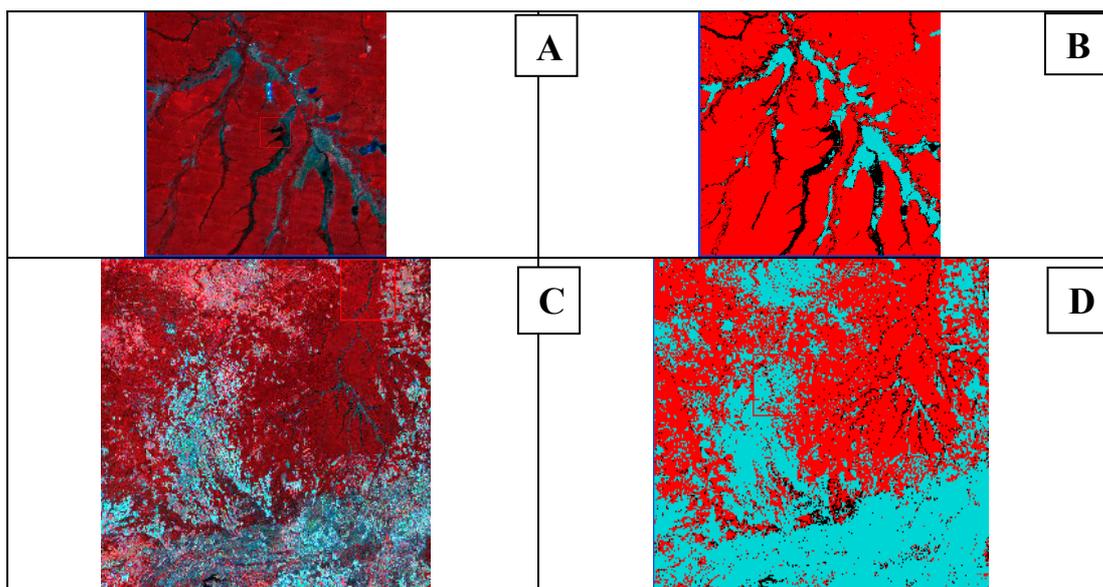


FIGURA 6 - Exemplo de comparação (análise visual) empregada entre: (A) Detalhes do produto MODIS; (B) detalhes do resultado obtidos pelo modelo - ênfase- água / áreas inundáveis; (C) o produto MODIS (dados 250m - MODIS) composição R (NIR) G (RED) e B (BLUE) do Estado do Mato Grosso e (D) o mapa final obtido com o modelo, classes: vermelho: floresta, cian – não floresta e preto – águas / áreas inundáveis.

A **Figura 8** mostra um esboço da interface criada após a verificação da viabilidade do processo metodológico criado. Sua elaboração partiu da necessidade de se criar uma maior interação junto ao usuário, tendo em vista que o processo se tornaria inviável com o emprego único e exclusivo de seus programas. Tanto a elaboração do modelo, bem como sua interação permitiu um ganho considerável na automação do processo metodológico e, sobretudo no tempo de processamento (aproximadamente duas horas) de suas etapas e obtenção do produto final.

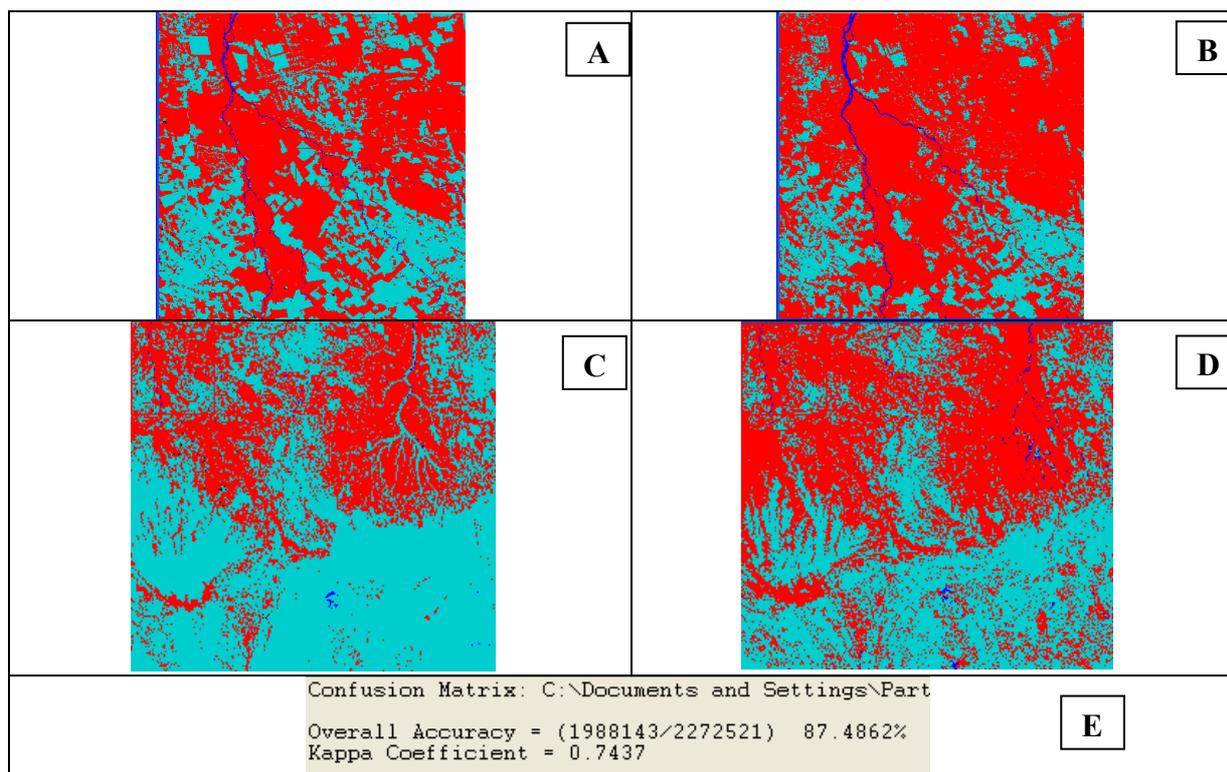


FIGURA 7 – Exemplo de melhor resultado obtido na comparação com os dados do Projeto PRODES (área do Estado do Mato Grosso – ano de 2002). (A) Detalhes do mapa final obtido com o Projeto PRODES; (B) detalhes do melhor resultado obtido com o modelo proposto (classes: floresta, não floresta e água/áreas inundáveis); e (C)/(D) seus respectivos Mapas finais. (E) Acurácia global e o coeficiente kappa do mapa final obtida empregando como verdade os dados do Projeto PRODES.

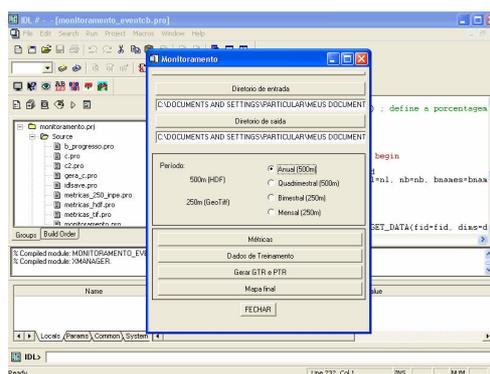


FIGURA 8 – Interface criada para a aplicação do sistema de monitoramento. Detalhes das etapas referentes à obtenção do produto final (seta em verde).

4. CONCLUSÕES

Com relação ao modelo proposto para monitoramento da cobertura terrestre pode-se concluir que: (1) é viável para a região amazônica, tendo em vista os resultados comparativos com os do Projeto PRODES; (2) permite o emprego de diversos tipos de dados de treinamento, tanto isoladamente, quanto em conjunto; e (3) tanto a elaboração do modelo, bem como sua interação permitiu um ganho considerável na automação do processo metodológico e,

sobretudo no tempo de processamento (aproximadamente duas horas) de suas etapas e obtenção do produto final.

Como forma de dar prosseguimento ao estudo realizado pode-se apresentar como sugestões: (1) criar novas métricas ou variáveis, implementá-las e testá-las para verificar a sua viabilidade no modelo proposto; (2) verificar a viabilidade do emprego dos novos produtos (cenas de alta resolução) na metodologia apresentada, como os gerados, futuramente, pelo SIVAM. Exemplos: HSS, MSS, entre outros; (3) ampliar ou aperfeiçoar a interface criada em função das novas necessidades e objetivos a serem criados pelo usuário final; e (4) verificar a possibilidade de migração de seu processo metodológico para a aplicação em programas livres.

Devido à necessidade de aplicação de modelos de monitoramento para as regiões tropicais, como o Brasil, a presente abordagem metodológica permite uma otimização no tratamento dos dados. Espera-se, que este tipo de produto gerado venha a contribuir para a nobre função da Força Aérea junto à Amazônia Brasileira, complementando, ainda, o já consagrado trabalho desenvolvido pelo INPE através do PRODES.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Anderson, L.O.; Shimabukuro, Y.E.; Latorre, M.L.; Espírito Santo, F.D.B.; Lima, A.; Duarte, V.; Arai, E.; Defries, R.S.; MORTON, D.; HANSEN, M.; JASINSKY, E. **Coleta de dados para estudos de cobertura arbórea, mapeamento da cobertura da Terra, desmatamento e avaliação de unidades de conservação/territórios indígenas: metodologias de trabalhos de campo**. São José dos Campos (INPE-12304-NTC/368), 2005, 36p.
- Breiman, L.; Friedman J.H.; Olshen R.A.; Stone C.J. **Classification and regression trees**. Belmont, CA: Wadsworth International, 1984. 358p.
- Defries, R.S.; Hansen, M.C.; Steining, M.; Derlayah, R.; Sohlberg R. & Townshend, J. 1997. Subpixel forest cover in Central Africa for multisensor, Multitemporal data. **Remote Sensing of Environment**, 60: 228-246
- Friedl, M.A.; Brodley, C.E. & Stralher, A.H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. **IEEE Transactions on Geoscience and Remote Sensing**, 37(2): 969-977.
- Hansen M.C.; Defries, R.S.; Townshend, J.R.G.; Carroll, M.; Dimiceli, C. And Sohlberg, A. global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm, **Earth Interactions**, 7(10):1-15, 2003.
- Hansen, M.C.; Defries, R.S.; Townshend, J.R.G.; Sohlberg, R.; Dimiceli, C.; Carroll, M. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. **Remote Sensing of Environment**, n.83, p.303-319, 2002.
- Hansen, M.C.; De Fries, R.S.; Townshend, J.R.G.; Sohlberg, R. Global land cover classification at 1km spatial resolution using a classification tree approach. **International Journal of Remote Sensing**, 21(6 e7):1331-1364, 2000.
- Reed, B.D.; Brown, J.F.; Vanderzee, D.; Loveland, T.R.; Merchant, J.W.; Ohlen, D.O.; Measuring phenological variability from satellite imagery. **Journal of Vegetation Science**, 5: 703-714, 1994.
- Strahler, A.; Muchoney, D.; Borak, J.; Friedl, M.; Gopal, S.; Lambin, E. & Moody, A. 1999. **MODIS Land Cover and Land-Cover Change products algorithm theoretical basis document (ATBD)**. Version 5.0, Center for Remote Sensing, Department of Geography, Boston University, May, 72p.
- Venables, W.W.; Ripley, B.D. **Modern applied statistics with S-PLUS**. Springer-Verlag, New York, 1994.