

Análise de Imagem Orientada a Objeto e Mineração de Dados aplicadas ao mapeamento geomorfológico e descoberta de conhecimento em mapas a partir de dados SRTM

Pedro José Farias Fernandes ^{1,2}

¹ Instituto Nacional de Pesquisas Espaciais – Divisão de Sensoriamento Remoto – DSR/INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
pjff@dsr.inpe.br

² Laboratório de Geografia Física – Universidade Federal Fluminense - LAGEF/UFF
Av. Gal. Milton Tavares de Souza s/nº - 24210-346 - Niterói - RJ, Brasil

Abstract. The objective is to contribute with a methodology for semi-automated regional geomorphological mapping data through Object Based Image Analysis, Data Mining and SRTM data, also generating knowledge for possible updates of old geomorphological maps at regional scale. This work uses the classes defined by the geomorphological map of the RADAMBRASIL Project. The RADAMBRASIL map was published in the scale of 1:1.000.000. The study area is located in Rio de Janeiro State and includes the Guanabara Bay. In general, the methodology involves five steps: (I) multiresolution segmentation of SRTM data, (II) selection of the training set, (III) data mining, (IV) classification, (V) manual edition of classification errors. This work obtained as results the decision trees, generated in the Data Mining step with a knowledge recovered from the original map, used in the geomorphological classification, and the refined maps for each taxonomic level, which were compared with the original classification. It can be concluded by visual analysis of these results that the adopted methodology can contribute for an efficient regional geomorphological mapping, and, also, can recover knowledge from old geomorphological maps, which can be used in the update, correction and refinement of these maps, especially when it is not possible contact the original author or specialist.

Palavras-chave: Geomorphology, OBIA, Data Mining, Geomorfologia, AIOO, Mineração de Dados.

1. Introdução

Modelos Digitais de Elevação (MDE) obtidos por Sensoriamento Remoto (SR) são um dos mais importantes dados utilizados em análise geoespacial (TOUTIN, 2004), uma vez que, com as técnicas computacionais disponíveis, favorecem a inclusão de dados topográficos em estudos dos sistemas terrestres e beneficiam os mapeamentos geomorfológicos. Entretanto, os mapas geomorfológicos antigos, como os do Projeto RADAMBRASIL envolveram técnicas de fotointerpretação que não contavam com os recursos computacionais disponíveis hoje em dia, bem como as bases de dados atuais (BARBOSA, 1973). Entretanto, esses mapas temáticos possuem importante conhecimento especializado inserido neles, pois o fotointérprete, ao delinear os polígonos de um mapa, faz uma análise das relações entre as classes e as condições ambientais da paisagem. Além disso, nem sempre é possível consultar os autores de um mapa para recuperar esse conhecimento, que pode servir para a atualização dos mapas (QI; ZHU, 2003).

As técnicas de processamento digital de imagens (PDI) são adequadas para a classificação geomorfológica a partir de MDE e de suas variáveis geomorfológicas (VG) derivadas. Entretanto, as classificações baseadas somente em *pixels* possuem limitações na definição de classes de relevo a partir de MDE por não incluírem relações topológicas e por tratarem os dados locais como entidades independentes (DRAGUT; BLASCHKE, 2006). Consequentemente, a classificação local (coloquialmente, *pixel a pixel*) de dados geomorfológicos gera um resultado de aspecto granuloso devido a sua variabilidade inerente e decorrente sobreposição dos valores entre as classes definidas por intervalos rígidos (DRAGUT, EISANK; 2012). Nesse sentido, a Análise de Imagem Orientada a Objeto (AIOO) trabalha com objetos gerados por um algoritmo de segmentação, que inclui, além de atributos texturais e estatísticos baseados nos valores locais, atributos de forma e topológicos (BENZ et al., 2004). A aplicação de AIOO ao mapeamento geomorfológico a partir de objetos

criados pelo uso do algoritmo de segmentação multiresolução sobre MDE é bem sucedida para esse propósito (DRAGUT; EISANK, 2012; DRAGUT; BLASCHKE, 2006). Porém, diante de inúmeros atributos disponíveis para a classificação, é necessário saber quais podem ser utilizados para caracterizar as classes com o mínimo de erro possível. Muitas vezes, esta percepção não é trivial, dada a multiplicidade de conhecimentos envolvidos e de possíveis padrões em um grande número de dados (HSU et al., 2002). Para tal problema, a Mineração de Dados (MD) envolve o uso de algoritmos para a compreensão dos dados e extração de padrões úteis na produção de um modelo de conhecimento (FAYYAD et al., 1996). Este modelo de conhecimento pode ser representado na forma de árvore de decisão (VIEIRA et al., 2012) e ser utilizado na classificação.

Portanto, admite-se como hipótese a existência de discontinuidades nos MDE e nas VG correspondentes às fronteiras reais no relevo, que podem viabilizar o mapeamento geomorfológico por AIOO e MD e modelar o conhecimento implícito em mapas geomorfológicos, como o do RADAMBRASIL. O objetivo deste estudo é contribuir com uma metodologia semiautomatizada para o mapeamento geomorfológico regional com dados SRTM através de AIOO e MD, gerando, também, conhecimento para possíveis atualizações de antigos mapeamentos geomorfológicos de escala regional.

2. Materiais e Métodos

2.1 Área de estudo

A área de estudo (Figura 1) está localizada nas Folhas SF.23/24 Rio de Janeiro/Vitória do Projeto RADAMBRASIL, no Estado do Rio de Janeiro, entre, aproximadamente, as latitudes -22° S e -23° S, e longitudes -43° W e -44° W. A área foi selecionada pela sua variedade topográfica e abriga seis Unidades Geomorfológicas mapeadas pelo RADAMBRASIL.

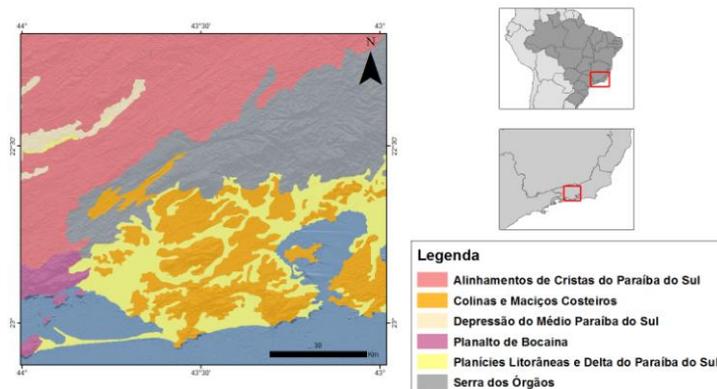


Figura 1. Localização da área de estudo e seu contexto geomorfológico.

2.2 Materiais e recursos de análise

Foram utilizados dados SRTM refinados e suas VG do projeto TOPODATA, cenas 22S45, 22S435, 23S45_ e 23S435, com *pixel* de 30 m: altimetria, declividade, curvaturas horizontal e vertical, e relevo sombreado. Duas imagens TM/Landsat-5 adquiridas no Catálogo de Imagens do INPE, cenas 217/76 e 217/75, banda 4, com resolução espacial de 30 m, foram utilizadas para a separação de Água e Continente no primeiro nível de segmentação. Também foram usados dois mapas geomorfológicos para eventuais consultas: (i) do Projeto RADAMBRASIL das Folhas SF.23/24 Rio de Janeiro/Vitória na escala 1:1.000.000 (RADAMBRASIL, 1983); (ii) do Projeto Rio de Janeiro do CPRM (CPRM, 2001) na escala 1:500.000 (CPRM, 2005). Para realizar o trabalho, os seguintes *softwares* foram utilizados: ArcGIS 9.3 (visualização de Planos de Informação (PIs) e confecção dos cartogramas); Definiens Developer 7.0 (segmentação e classificação); ENVI 4.7 (processamento digital de imagens); WEKA 3.6 (mineração de dados).

3.3 Métodos

O fluxograma da Figura 2 mostra a sequência dos procedimentos adotados no trabalho:

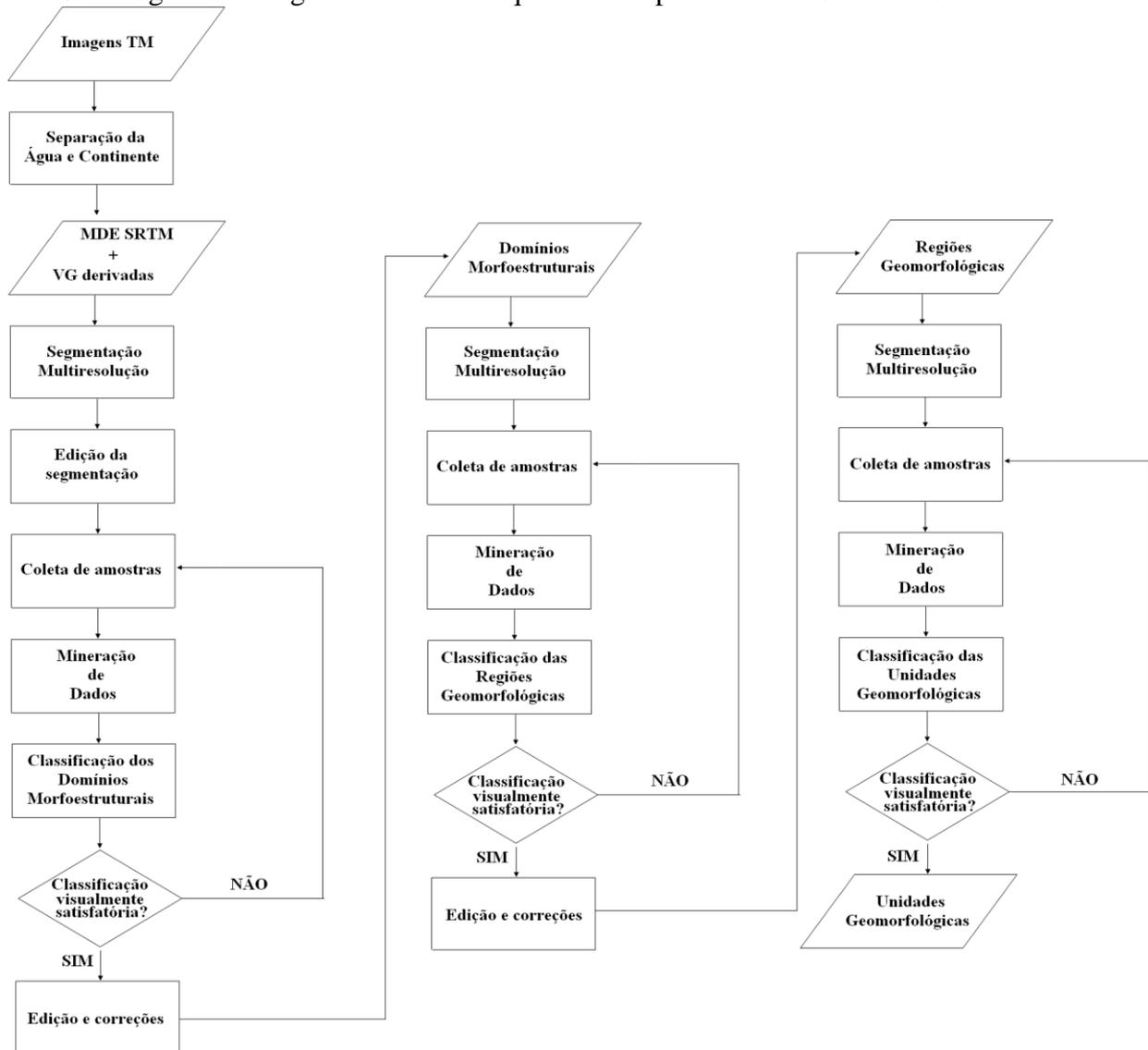


Figura 2. Fluxograma das etapas adotadas no trabalho.

3.3.1 Definição das classes geomorfológicas

As classes geomorfológicas mapeadas pelo Projeto RADAMBRASIL estão distribuídas em 3 táxons e foram utilizadas neste trabalho (Figura 3). O 1º táxon refere-se aos Domínios Morfoestruturais, o 2º às Regiões Geomorfológicas e o 3º às Unidades Geomorfológicas.

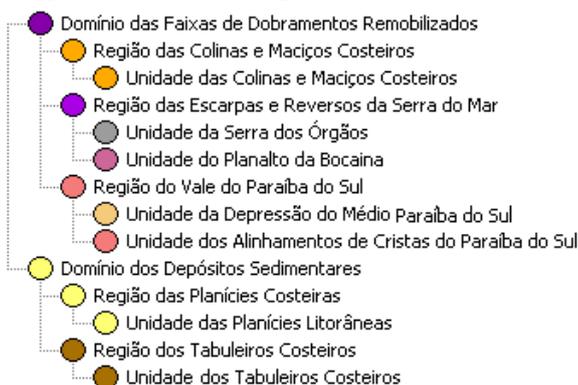


Figura 3. Classes geomorfológicas existentes na área de estudo por nível taxonômico

3.3.2 Segmentação Multiresolução

O primeiro passo da AIOO é a segmentação, que cria os objetos a serem classificados. Os objetos são um pré-requisito para a AIOO (BAATZ; SCHÄPE, 2000) e constituem a unidade básica desse tipo de processamento (BENZ et al., 2004). A segmentação foi realizada pelo algoritmo de segmentação multiresolução (BAATZ; SCHÄPE, 2000), que individualiza polígonos, tratados como objetos, a partir de dados espacializados, em vários níveis de segmentação. Esse algoritmo baseia-se no crescimento de regiões e é configurado com base em três parâmetros: (i) Escala (*scale*): medida de heterogeneidade que define o tamanho dos objetos; (ii) Formato (*shape*): referente à participação da forma na segmentação; um valor baixo implica em maior peso para os atributos locais; e (iii) Compacidade (*compactness*): um alto valor tende a produzir objetos compactos, um valor baixo cria objetos com bordas mais suaves. Os objetos são evidenciados em diferentes escalas (BAATZ; SCHÄPE, 2000), e são confrontados a seus vizinhos, seus superobjetos e subobjetos (BENZ et al., 2004). Foi utilizada uma segmentação a partir do primeiro nível de segmentação (de separação entre Água e Continente) até o nível referente às Unidades Geomorfológicas, numa abordagem referida como *top-down*. Assim se obedece à hierarquia taxonômica tradicional, conforme corroborado pela recomendação de Ross (1992), de que a classificação geomorfológica deva ter os Domínios Morfoestruturais como ponto de partida, por fornecerem as características principais do relevo mapeado. Desse modo, o segundo nível de segmentação foi utilizado para classificar os Domínios Morfoestruturais, o terceiro na classificação das Regiões Geomorfológicas, e o quarto nível utilizado no mapeamento das Unidades Geomorfológicas.

Primeiro, os objetos foram classificados, no primeiro nível de segmentação, em duas classes: Água (objetos com Média da Banda 4 \leq 48,1) e Continente (objetos que não foram classificados como Água). Após a classificação automática, foram efetuadas edições manuais, para obter a classificação final. Finalmente, os objetos classificados foram fusionados em um só objeto por classe, e este nível de segmentação foi copiado para o nível abaixo.

Como visto na Figura 2, a classe Continente foi segmentada e originou objetos referentes aos Domínios Morfoestruturais. Esses, por sua vez, após a classificação pela árvore de decisão gerada na MD, foram unidos em um único objeto por classe, e foram segmentados novamente no nível abaixo para criar os objetos das Regiões Geomorfológicas. Esse procedimento foi repetido a partir das Regiões Geomorfológicas, classificadas após a MD, para o mapeamento das Unidades Geomorfológicas no quarto nível de segmentação. Entretanto, houve alguns erros na segmentação do segundo nível (Domínios Morfoestruturais), por causa de perturbações nos dados, como em áreas costeiras planas onde a urbanização é densa. A segmentação originou, nessas áreas, objetos com médias elevadas de declividade. Portanto, foi feita a edição manual da segmentação para o fornecimento correto de amostras de classes para a MD. Depois, foram feitos testes para definir a segmentação ideal por classe. As configurações das segmentações por classe estão na Tabela 1 abaixo.

Ressalta-se que as classes Colinas e Maciços Costeiros, Planícies Costeiras e Tabuleiros Costeiros (no terceiro nível de segmentação) permanecem iguais no quarto nível, portanto não houve necessidade de segmentação dessas classes e da etapa de MD para a posterior classificação. Além disso, a segmentação do quarto nível para a separação do Planalto da Bocaina da Serra dos Órgãos (Tabela 1) criou um único objeto da Unidade do Planalto da Bocaina. Entretanto, para a MD, é preciso fornecer mais de uma amostra por classe. Por esse motivo, foi necessário criar mais objetos referentes ao Planalto da Bocaina para fornecer mais amostras na MD. Dessa forma, a escala da segmentação foi reduzida de 1249 para 468.

Tabela 1. Configurações da Segmentação Multiresolução para as classes

Nível	Classes	Variável	Compac.	Formato	Escala
Primeiro	Água e Continente	Banda 4	0,3	0,3	50
Segundo	Faixas de Dobramentos Remobilizados e Depósitos Sedimentares	Declividade	0,1	0,1	85
Terceiro	Colinas e Maciços Costeiros, Escarpas e Reversos da Serra do Mar e Vale do Paraíba do Sul	Declividade	0,4	0,5	268
	Planícies Costeiras e Tabuleiros Costeiros	Declividade	0,1	0,1	131
Quarto	Planalto da Bocaina e Serra dos Órgãos	Declividade	0,5	0,9	1249
	Alinhamentos de Cristas do Paraíba do Sul e Depressão do Médio Paraíba do Sul	Altimetria	0,7	0,4	953

3.3.2 Estabelecimento do conjunto de treinamento e Mineração de Dados

Objetos foram selecionados como amostras das classes para a MD. Entretanto, a MD não foi feita de uma só vez, e sim em etapas, para a redução de erros de classificação. No segundo nível, o conjunto usado possui 38 amostras de Depósitos Sedimentares e 142 de Faixas de Dobramentos Remobilizados. No terceiro, foram 40 amostras de Colinas e Maciços Costeiros, 20 de Escarpas e Reversos da Serra do Mar e 20 amostras do Vale do Paraíba do Sul. No quarto nível de segmentação, todos os objetos de cada classe foram utilizados na MD, que foi feita duas etapas. A primeira etapa foi feita para a definição de árvores de decisão que diferenciavam o Planalto da Bocaina da Serra dos Órgãos. A segunda foi empregada para definir a separação entre os Alinhamentos de Cristas do Paraíba do Sul e a Depressão do Médio Paraíba do Sul. No total foram 3 amostras do Planalto da Bocaina, 20 da Serra dos Órgãos, 30 de Alinhamentos de Cristas do Paraíba do Sul e 4 da Depressão do Médio Paraíba do Sul. As amostras foram exportadas em tabelas com as classes e seus atributos para a MD.

Com exceção dos Domínios Morfoestruturais, onde a MD foi feita com base em atributos geomorfométricos apenas, decidiu-se fazer um teste para verificar o impacto do uso de atributos de localização na mineração de dados e na classificação geomorfológica. Portanto, foram feitos dois experimentos: (a): apenas atributos geomorfométricos foram exportados: as médias da altimetria, da declividade, das curvaturas (vertical e horizontal), e o desvio padrão destas; e (b): além dos atributos geomorfométricos, foram exportados dois atributos de localização que possuem potencial para separar as classes: distância do Oceano e distância da Planície do Paraíba do Sul. A MD foi feita pelo algoritmo C4.5 (disponível no WEKA como J48), devido aos bons resultados obtidos em SR por Vieira et al. (2012) e outros.

3.3.2 Classificação e edição

As árvores de decisão geradas pela MD foram construídas no Definiens na forma de rede hierárquica. Para respeitar a estrutura das árvores, a relação entre classes foi feita pela definição de regras booleanas, como feito por Vieira (2012). Depois, foram feitas as classificações geomorfológicas. Os erros de classificação foram corrigidos pela edição manual da segmentação, feita com base na interpretação visual do relevo sombreado. No caso da classe Tabuleiros Costeiros (terceiro nível de segmentação), devido a sua pequena área de extensão, foi decidido não utilizar a MD na sua classificação, que foi feita de acordo com seu valor exato de declividade, diretamente. Como o objeto criado de Tabuleiros Costeiros possui uma média de declividade igual a 5,735, esse valor foi inserido na descrição dessa classe.

4. Resultados

4.1 Conhecimento extraído do mapa RADAMBRASIL na forma de árvores de decisão

As árvores de decisão representam um modelo de conhecimento extraído do mapa original RADAMBRASIL (figura 4).

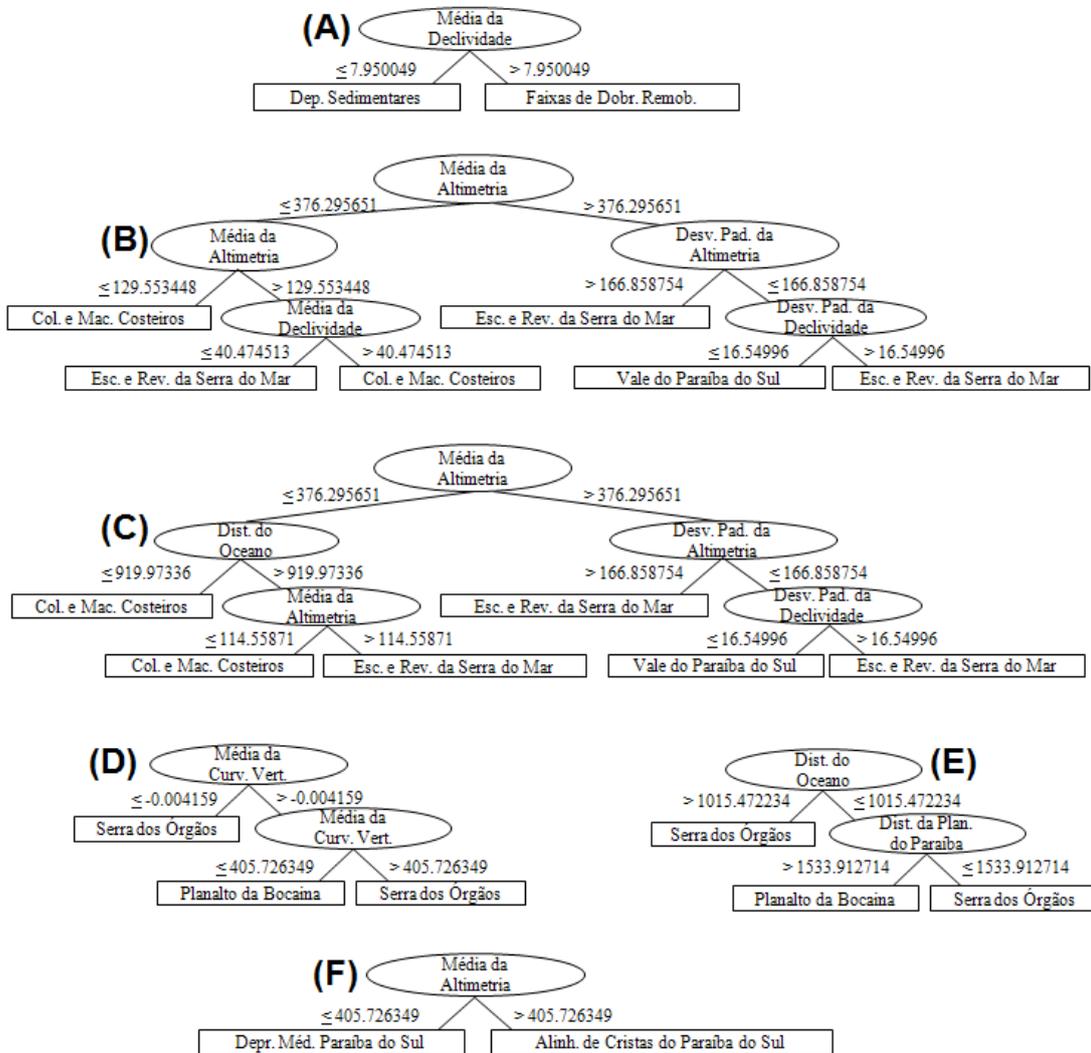


Figura 4. Árvores de decisão geradas na mineração de dados. (A) Árvore dos Domínios Morfoestruturais. (B) Árvore das Regiões Geomorfológicas somente com VG. (C) Árvore das Regiões Geomorfológicas com VG e atributos de localização. (D) Árvore que separa Planalto da Bocaina e Serra dos Órgãos somente com VG. (E) Árvore que separa Planalto da Bocaina e Serra dos Órgãos com VG e atributos de localização. (F) Árvore que separa a Depressão do Médio Paraíba do Sul dos Alinhamentos de Cristas do Paraíba do Sul.

Com exceção da árvore que separa os Alinhamentos de Cristas do Paraíba do Sul da classe Depressão do Médio Paraíba do Sul, quando os atributos de localização foram incluídos na MD, estes foram selecionados pelo algoritmo J48. Além disso, os dois tipos de árvores de decisão (com e sem atributos de localização) foram testados nas classificações. As classificações feitas com árvores que incluem atributos de localização geraram resultados que, visualmente, apresentaram menor quantidade de objetos classificados incorretamente. Isso sugere alguma conformidade com os resultados obtidos por Qi e Zu (2003), já que esses autores mostram que a exatidão das árvores de decisão obtidas a partir de amostras de classes de solos de um mapa pedológico aumenta com a inclusão de informação topológica em um conjunto de atributos geomorfométricos, geradas na etapa de MD. Portanto, foram utilizadas

no prosseguimento do trabalho as árvores que forneceram classificações com menores quantidades de erros no Definiens, ou seja, árvores que incluem atributos de localização. Entretanto, pelo desempenho fraco da MD na separação do Planalto da Bocaina e da Serra dos Órgãos, e como a segmentação descrita na Tabela 1 delimitou bem o Planalto da Bocaina, utilizou-se o conhecimento especialista na classificação: o objeto da classe de Escarpas e Reversos da Serra do Mar que faz fronteira com a classe Água foi classificado como Planalto da Bocaina, o restante como Serra dos Órgãos.

4.2 Mapas geomorfológicos

É possível notar que a metodologia forneceu resultados com limites refinados em relação ao desenho original do RADAMBRASIL (Figura 5). A segmentação do MDE forneceu um resultado mais objetivo e envolveu o uso de técnicas computacionais, que facilitaram a uniformidade de critérios. Além disso, os dados SRTM estão em uma escala maior que a do mapa original, o que permite o refinamento desses limites. Outro fato notório é a possibilidade de se alterar, redesenhar, atualizar e até mesmo corrigir classes do mapa original. Por exemplo, a partir do segundo nível taxonômico (terceiro de segmentação) é possível notar que as ilhas existentes no 3º quadrante do mapa original foram classificadas como Escarpas e Reversos da Serra do Mar devido a uma semelhança litológica (RADAMBRASIL, 1983). Entretanto, com base em consultas ao mapa do CPRM, foi decidido incluí-las na Região das Colinas e Maciços Costeiros. Também, nota-se que a metodologia permitiu a classificação da planície do Vale do Paraíba do Sul, no 1º e 2º quadrantes dos mapas na direção SO-NE, como um Depósito Sedimentar.

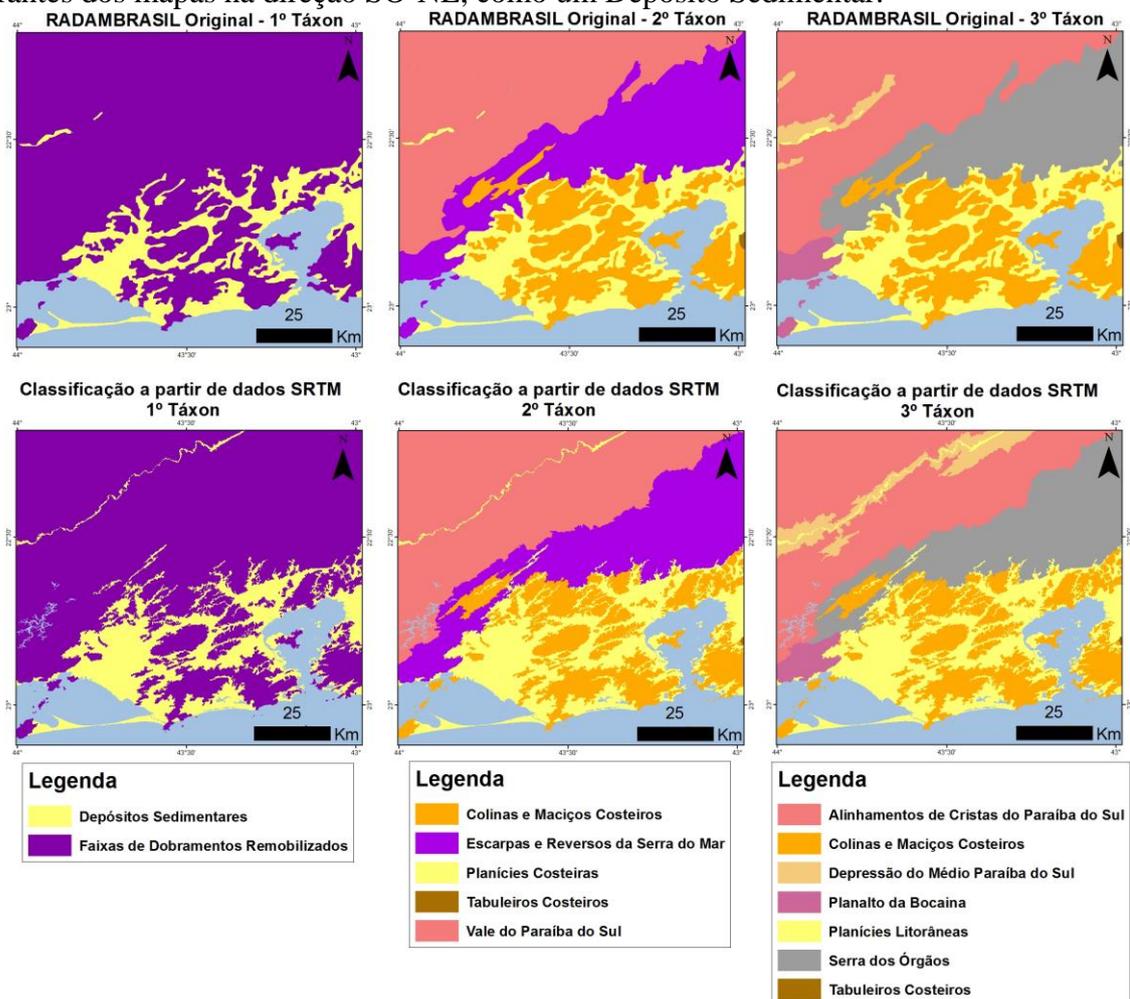


Figura 5. Comparação entre os mapas originais e a classificação dos dados SRTM.

5. Conclusão

Os resultados permitiram a confirmação da hipótese do presente trabalho. Há descontinuidades nos dados SRTM que refletem mudanças no relevo. A segmentação delimitou os diferentes tipos de relevo na área de estudo e criou os objetos necessários para a classificação de acordo com as classes estabelecidas pelo RADAMBRASIL.

A MD tornou possível a seleção dos atributos úteis para a classificação geomorfológica. Isso tornou a metodologia com um caráter mais automatizado. Entretanto, seu uso requer alguns cuidados. É importante utilizar apenas os atributos com potencial de separar as classes, e aplicar a MD várias vezes na separação de poucas classes geomorfológicas, para evitar a geração de árvores de decisão muito complexas e que confundem as classes.

Visualmente, percebe-se que as ferramentas utilizadas permitiram a geração de novos mapas que estão de acordo com a classificação do RADAMBRASIL, e que possuem limites refinados por conta do uso dos dados SRTM. Além disso, a metodologia utilizada permitiu a extração de um modelo de conhecimento implícito no mapa RADAMBRASIL. Esse conhecimento pode ser utilizado para atualização e refinamento do mapa, pois não é sempre possível contar com a disponibilidade dos autores antigos para realizar tal tarefa.

6. Bibliografia

- Baatz, M.; Schäpe, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. (eds.). *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT Symposium Salzburg 2000*. Karlsruhe. Herbert Wichmann Verlag, 2000. p. 12-23.
- Barbosa, G. V. Cartografia geomorfológica utilizada pelo projeto RADAM. In: *CONGRESSO BRASILEIRO DE GEOLOGIA*, 27., Aracaju, 1973, *Anais...* Aracaju, SBG, 1973, v.1, p. 427-433.
- Benz, U. C.; Hofmann, P.; Wikkhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *Photogrammetry & Remote Sensing*, v. 58, p. 239-258, 2004.
- CPRM. **Projeto Rio de Janeiro**. Brasília: Ministério de Minas e Energia, 2001, 614 p.
- Dragut, L.; Blaschke, T. Automated classification of landform elements using object-based image analysis. *Geomorphology*, v. 81, p. 330-344, 2006.
- Dragut, L.; Eisank, C. Automated object-based classification of topography from SRTM data. *Geomorphology*, v. 141-142, n. 1, p. 21-33, 2012.
- Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. *Advances in knowledge discovery and Data mining*. Cambridge: MIT Press, 1996.
- Hsu, W.; Lee, M. L.; Zhang, J. Image mining: trends and developments. *Journal of Intelligent Information Systems*, v. 19, n. 1, p. 7-23, 2002.
- Qi, F.; Zhu A. Knowledge Discovery from soil maps using inductive learning. *International Journal of Geographical Information Science*, v. 17, n. 8, p.771-795, 2003.
- Ross, J. L. S. O registro cartográfico dos fatos geomórficos e a questão da taxonomia do relevo. *Revista do Departamento de Geografia*, n. 6, p. 17 - 29, 1992.
- Toutin, T. RADARSAT-2 stereoscopy and polarimetry for 3D mapping. *Canadian Journal of Remote Sensing*, v. 30, n. 3. p. 496-503, 2004.
- Vieira, M.; Formaggio, A. R.; Rennó, C. D.; Atzberger, C.; Aguiar, D. A.; Mello, M. P. Object based image analysis and data mining applied to a remotely sensed Landsat time-series to map sugarcane over large áreas. *Remote Sensing of Environment*, v.123, p. 533-562, 2012.