

UNIVERSIDADE FEDERAL DE ALFENAS – MG

RODRIGO CESÁRIO JUSTINO

**MAPEAMENTO DO USO DA TERRA E DA COBERTURA VEGETAL DA SUB-
BACIA HIDROGRÁFICA DO RIO SÃO TOMÉ, SUL DE MINAS GERAIS, POR
MINERAÇÃO DE DADOS UTILIZANDO IMAGENS IRS-P6/ LISS III**

Alfenas/MG
2014

RODRIGO CESÁRIO JUSTINO

**MAPEAMENTO DO USO DA TERRA E DA COBERTURA VEGETAL DA SUB-
BACIA HIDROGRÁFICA DO RIO SÃO TOMÉ, SUL DE MINAS GERAIS, POR
MINERAÇÃO DE DADOS UTILIZANDO IMAGENS IRS-P6/ LISS III**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ecologia e Tecnologia Ambiental pela Universidade Federal de Alfenas/MG. Área de concentração: **Meio Ambiente, Sociedade e Diversidade Biológica**. Orientador: Prof. Dr. Fernando Shinji Kawakubo.

Alfenas/MG
2014

RODRIGO CESÁRIO JUSTINO

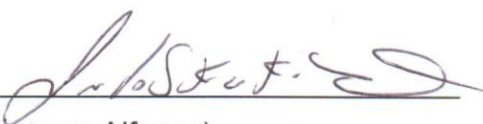
MAPEAMENTO DO USO DA TERRA E DA COBERTURA VEGETAL DA SUB-BACIA HIDROGRÁFICA DO RIO SÃO TOMÉ, SUL DE MINAS GERAIS, POR MINERAÇÃO DE DADOS UTILIZANDO IMAGENS IRS-P6/ LISS III

A Banca examinadora abaixo, aprova a Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ecologia e Tecnologia Ambiental pela Universidade Federal de Alfenas/MG. Área de concentração: **Meio Ambiente, Sociedade e Diversidade Biológica.**

Aprovado em: 29/07/2014.

Profº. Dr. Fernando Shinji Kawakubo

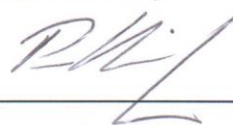
Assinatura:



Instituição: Universidade Federal de Alfenas/MG (Campus Alfenas)

Profº. Dr. Paulo Henrique de Souza

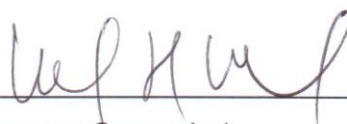
Assinatura:



Instituição: Universidade Federal de Alfenas/MG (Campus Alfenas)

Profª. Dr. Marcos Roberto Martines

Assinatura:



Instituição: Universidade Federal de São Carlos/SP (Campus Sorocaba)

Dedico a Deus, aos meus pais Elaine e Pedro e familiares e aos meus amigos pelo apoio na realização deste trabalho.

AGRADECIMENTOS

Ao Programa de Pós-Graduação em Ecologia e Tecnologia Ambiental da Universidade Federal de Alfenas/MG pela oportunidade oferecida.

Ao Prof. Dr. Fernando Shinji Kawakubo, orientador, pela dedicação, orientação e auxílio ao longo de todo o processo de realização deste trabalho.

Aos Profs. Dr. Marcos Roberto Martines e Dr. Paulo Henrique de Souza pela participação como membros titulares da minha banca de defesa.

Ao Prof. Dr. Ronaldo Luiz Mincato e à Profa. Dra. Rúbia Gomes Morato pela participação em minha banca de qualificação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Ao corpo docente do Programa de Pós-Graduação em Ecologia e Tecnologia Ambiental da Universidade Federal de Alfenas/MG.

Ao Laboratório de Geoprocessamento do curso de Geografia da Universidade Federal de Alfenas/MG.

Ao Dr. Thales Sehn Korting pelos esclarecimentos com relação ao uso do aplicativo GeoDMA.

A todos os meus familiares e amigos que me apoiaram nesta empreitada.

RESUMO

O objetivo deste estudo foi realizar uma classificação do uso da terra e cobertura vegetal na sub-bacia do rio São Tomé, no sul do Estado de Minas Gerais, utilizando imagens multiespectrais geradas pelo sensor LISS III (*Linear Imaging Self-Scanner*) a bordo do satélite IRS (*Indian Resource Satellite*) e técnicas de mineração de dados. A área de estudo localiza-se em uma região de grande tradição de cultivo de café cujas propriedades são predominantemente de pequeno e médio portes. Para realizar a classificação das imagens, foi utilizado o aplicativo GeoDMA (*Geographic Data Mining Analyst*) que possui algoritmos para segmentação, extração de atributos, seleção de feições e classificação. Os resultados mostraram o grande potencial das técnicas de mineração de dados na classificação de imagens de satélite. Algumas das vantagens da utilização de mineração de dados incluem: a possibilidade de incorporar no processo de classificação um grande número de variáveis, sejam espectrais, espaciais e atributos do terreno; a geração de classificações consistentes e a simplicidade para interpretar a suas estruturas de classificação.

PALAVRAS-CHAVE: GeoDMA, mineração de dados, sensoriamento remoto, segmentação, atributos espectrais, atributos espaciais, árvore de decisão, exatidão.

ABSTRACT

The objective of this study was to perform a classification of land use and land cover of the São Tomé river watershed located in the south of the Minas Gerais State by using multispectral imagery collected by the Linear Imaging Self-Scanner onboard the Indian Resource Satellite and data mining techniques. The study area lies in a region with great tradition of coffee cultivation whose properties are predominantly of small and medium sizes. To carry out the image classification, we used the Geographic Data Mining Analyst (GeoDMA), a toolbox that has algorithms for segmentation, feature extraction, feature selection and classification. The results showed the great potential of the data mining techniques for remote sensing imagery classification. Some advantages encompass the possibility of incorporate in the classification process a great variety of informations that includes spectral, spatial and topographic attributes, the generation of high classification accuracy, and the simplicity to interpret their classification structure.

KEYWORDS: GeoDMA, data mining, remote sensing, segmentation, spectral attributes, spatial attributes, decision tree, accuracy.

LISTA DE QUADROS

Quadro 1 - Fórmulas preexistentes no ILWIS.....	25
Quadro 2 - Características espaciais baseadas em segmentação.....	51

LISTA DE TABELAS

Tabela 1 - Parâmetros cartográficos definidos no projeto.....	23
Tabela 2 - Características das bandas da imagem IRS-P6/ LISS III.....	31
Tabela 3 - Diferentes valores para compacidade, escala e cor.....	43
Tabela 4 - Matriz de confusão.....	48
Tabela 5 - Valores do índice Kappa e o seu grau de concordância.....	49

LISTA DE FIGURAS

Figura 1 - Mapa de localização da sub-bacia do rio São Tomé.....	18
Figura 2 - Fluxograma dos passos da metodologia.....	20
Figura 3 - Área de delimitação da sub-bacia.....	22
Figura 4 - Modelo de superfície com os eixos x, y e z.....	23
Figura 5 - Cálculo da menor distância entre os contornos.....	24
Figura 6 - Mapa de hipsometria.....	27
Figura 7 - Mapa de declividade.....	28
Figura 8 - Mapa de orientação das vertentes.....	29
Figura 9 - Mapa de ordem hierárquica dos cursos d'água.....	30
Figura 10 - Os três interpoladores presentes no ILWIS.....	33
Figura 11 - Exemplo de segmentação por crescimento de regiões.....	34
Figura 12 - Interação do GeoDMA como <i>plugin</i> do TerraView.....	35
Figura 13 - Mudança na árvore de decisão.....	39
Figura 14 - Estrutura e aparência de uma árvore de decisão.....	40
Figura 15 - Precisão e exatidão.....	41
Figura 16 - Pontos aleatórios de verdades terrestres.....	42
Figura 17 - Resultados da segmentação utilizando diferentes parâmetros.....	43
Figura 18 - Árvore de decisão para a classificação por mineração de dados....	44
Figura 19 - Mapa final de uso da terra e cobertura vegetal.....	47

SUMÁRIO

	RESUMO	11
	ABSTRACT	12
1	INTRODUÇÃO	13
2	OBJETIVOS	16
3	JUSTIFICATIVA	16
4	MATERIAL E MÉTODOS	17
4.1	ÁREA DE ESTUDO.....	17
4.2	ESTRUTURAÇÃO DA BASE CARTOGRÁFICA E CRIAÇÃO DE UM PROJETO EM SIG.....	21
4.3	DERIVAÇÕES DE INFORMAÇÕES DO TERRENO.....	23
4.4	PROCESSAMENTO DAS IMAGENS.....	31
4.4.1	IMAGENS LISS III.....	31
4.4.2	CORREÇÃO GEOMÉTRICA.....	31
4.5	SEGMENTAÇÃO E CLASSIFICAÇÃO.....	33
4.5.1	SEGMENTAÇÃO.....	33
4.5.2	CLASSIFICAÇÃO POR MINERAÇÃO DE DADOS.....	37
4.5.3	ALGORITMO C4.5.....	39
4.5.4	ANÁLISE DE EXATIDÃO.....	40
5	RESULTADOS	43
6	DISCUSSÃO	50
7	CONCLUSÃO	53
	AGRADECIMENTOS	54
	REFERÊNCIAS BIBLIOGRÁFICAS	54
	APÊNDICE A – ARTIGO SUBMETIDO	61

MAPEAMENTO DO USO DA TERRA E DA COBERTURA VEGETAL DA SUB-BACIA HIDROGRÁFICA DO RIO SÃO TOMÉ, SUL DE MINAS GERAIS, POR MINERAÇÃO DE DADOS UTILIZANDO IMAGENS IRS-P6/ LISS III

Rodrigo Cesário Justino¹
Fernando Shinji Kawakubo²

*¹Programa de Pós-Graduação em Ecologia e Tecnologia Ambiental (Mestrado)/
Universidade Federal de Alfenas – UNIFAL-MG*

*²Docente do curso de Geografia/ Instituto de Ciências da Natureza/ Universidade
Federal de Alfenas – UNIFAL-MG*

[1rodrigofriend@bol.com.br](mailto:rodrigofriend@bol.com.br), [2fskgeo@gmail.com](mailto:fskgeo@gmail.com)

RESUMO

O objetivo deste estudo foi realizar uma classificação do uso da terra e cobertura vegetal na sub-bacia do rio São Tomé, no sul do Estado de Minas Gerais, utilizando imagens multiespectrais geradas pelo sensor LISS III (*Linear Imaging Self-Scanner*) a bordo do satélite IRS (*Indian Resource Satellite*) e técnicas de mineração de dados. A área de estudo localiza-se em uma região de grande tradição de cultivo de café cujas propriedades são predominantemente de pequeno e médio portes. Para realizar a classificação das imagens, foi utilizado o aplicativo GeoDMA (*Geographic Data Mining Analyst*) que possui algoritmos para segmentação, extração de atributos, seleção de feições e classificação. Os resultados mostraram o grande potencial das técnicas de mineração de dados na classificação de imagens de satélite. Algumas das vantagens da utilização de mineração de dados incluem: a possibilidade de incorporar no processo de classificação um grande número de variáveis, sejam espectrais, espaciais e atributos do terreno; a geração de classificações consistentes e a simplicidade para interpretar a suas estruturas de classificação.

PALAVRAS-CHAVE: GeoDMA, mineração de dados, sensoriamento remoto, segmentação, atributos espectrais, atributos espaciais, árvore de decisão, exatidão.

ABSTRACT

The objective of this study was to perform a classification of land use and land cover of the São Tomé river watershed located in the south of the Minas Gerais State by using multispectral imagery collected by the Linear Imaging Self-Scanner onboard the Indian Resource Satellite and data mining techniques. The study area lies in a region with great tradition of coffee cultivation whose properties are predominantly of small and medium sizes. To carry out the image classification, we used the Geographic Data Mining Analyst (GeoDMA), a toolbox that has algorithms for segmentation, feature extraction, feature selection and classification. The results showed the great potential of the data mining techniques for remote sensing imagery classification. Some advantages encompass the possibility of incorporate in the classification process a great variety of informations that includes spectral, spatial and topographic attributes, the generation of high classification accuracy, and the simplicity to interpret their classification structure.

KEYWORDS: GeoDMA, data mining, remote sensing, segmentation, spectral attributes, spatial attributes, decision tree, accuracy.

1. INTRODUÇÃO

No intuito de criar condições para a sua sobrevivência, o homem, ao longo de sua história, tem se apropriado da paisagem para a construção de moradias, para produzir alimentos e para gerar riquezas através da exploração dos recursos naturais, da criação de animais em pastos e das grandes plantações de monocultura. Entretanto, antes do desenvolvimento da agricultura, o homem era nômade e na sua forma primitiva de subsistência sobrevivia à base da caça e da coleta de alimentos. Com o passar do tempo, o homem foi se tornando menos nômade e mais dependente da terra em que vivia. Famílias, clãs e vilas se desenvolveram e, com isto, o desenvolvimento da habilidade de produzir, ou seja, surgiu a agricultura (LOPES; GUILHERME, 2007). A partir daí é que surgiram condições para a sua fixação e tornou-se possível o desenvolvimento das civilizações.

Os primeiros sistemas de cultivo e de criação apareceram no período neolítico, há menos de dez mil anos e daí se expandiu pelo mundo de duas formas principais; pelos sistemas pastorais em regiões com vegetação herbácea (estepes e savanas) e pelo cultivo de derrubada-queimada na maior parte das zonas de florestas temperadas e tropicais (MAZOYER; ROUDART, 2010).

No Brasil as primeiras experiências de cultivo de plantas teriam ocorrido no Planalto Central há aproximadamente 6 mil anos, quando se firmaram os climas tropical e subtropical (LOPEZ; MOTA, 2008). Já a ocupação econômica das terras brasileiras pelos portugueses se deu segundo Furtado (2005), em consequência da pressão política exercida sobre Portugal e Espanha pelas demais nações europeias. Ao longo de sua história, o Brasil passou por inúmeros ciclos econômicos importantes, tais como o ciclo do ouro (século XVII-XVIII), o ciclo da borracha (século XIX-XX) e os ciclos econômicos atrelados à agricultura, se destacando o da cana-de-açúcar (século XVI-XVII), do café (século XVIII-XIX) e do cacau (século XIX-XX).

A agricultura exerceu e continua exercendo papel fundamental no desenvolvimento econômico do país. Atualmente, o Brasil possui uma grande diversidade em termos de produção agrícola. Esta diversidade se deve, em grande parte, à forma como o território foi ocupado, a dimensão continental e diversidade climática e cultural de suas regiões (ROSS, 2005).

Não obstante a importância que possui na economia brasileira, a agricultura muitas vezes é praticada sem um planejamento adequado, sendo que a retirada da vegetação nativa ou a substituição por outro tipo de uso da terra tem agravado o processo da fragmentação florestal e provocado consequências negativas nos diferentes compartimentos da natureza, afetando inúmeras espécies da fauna e da flora (SOARES et al., 2011), impactando ainda neste processo o bem estar da própria sociedade.

O planejamento agrícola e ambiental é, portanto, de suma importância para minimizar os possíveis impactos negativos que a intervenção humana pode trazer ao meio ambiente e à própria sociedade. Pessoa; Fernandes (2010) citam que o planejamento vem sendo apontado por vários autores como o ponto-chave para a integração harmoniosa entre o desenvolvimento e a necessidade de proteção e conservação do meio ambiente.

Estudos de capacidade de uso da terra e de aptidão agrícola (IBGE, 2006), bem como da fragilidade ambiental do terreno que compromete a funcionalidade do sistema, quebrando o seu estado de equilíbrio dinâmico (SPÖRL; ROSS, 2004), são instrumentos valiosos na elaboração do planejamento agrícola e ambiental, pois definem o potencial do solo para o cultivo e as limitações físicas do terreno com relação ao desencadeamento de processos erosivos.

A tecnologia do sensoriamento remoto (SR) constitui em uma das principais fontes de informações de apoio aos estudos destinados ao planejamento agrícola e ambiental, pois, além de fornecer uma visão geral da paisagem e registrar comprimentos de ondas inacessíveis à visão humana (MACHADO; QUINTANILHA, 2008), o SR também permite obter informações em intervalos tempo curtos entre uma data e outra e a um baixo custo na aquisição (KAWAKUBO, 2010).

O Brasil possui grande tradição na utilização dos recursos de SR para a geração de informação territorial. Desde a década de 1970, o INPE (Instituto Nacional de Pesquisas Espaciais) recebe imagens da série dos satélites Landsat (*Land Remote Sensing Satellite*) em uma estação de recepção própria localizada em Cuiabá-MT, centro geográfico da América do Sul (FLORENZANO, 2008). As imagens Landsat permitem em conjunto, quando analisadas, reconstituir a história de uso e ocupação do território brasileiro nas últimas quatro décadas. Estas imagens são armazenadas em um arquivo que conta também com um volumoso acervo de

imagens geradas pelos satélites da série CBERS (*China Brazil Earth Resources Satellite*) e mais recentemente do IRS (*Indian Resources Satellite*).

Até há poucos anos, a interpretação das imagens de satélite para o mapeamento do uso da terra e cobertura vegetal era feita de maneira analógica, utilizando como recurso os elementos de reconhecimento da fotointerpretação herdados das fotografias aéreas (CERON; DINIZ, 1966).

Já as técnicas de processamento de imagens voltadas para o reconhecimento de padrões, por sua vez, adotam critérios objetivos no reconhecimento das classes. Apesar de muitas vezes as classificações digitais não atingirem os níveis de exatidão exigidos, o método permite economia de tempo na execução do mapeamento e os produtos gerados podem ser facilmente integrados a uma base de informações gerenciada por um Sistema de Informação Geográfica (SIG).

Vários trabalhos têm sido desenvolvidos no território nacional, fazendo uso dos recursos do SR e das técnicas de processamento digital de imagens (PDI) para analisar e mapear o uso da terra e cobertura vegetal em diversos níveis de enfoque que variam da escala regional (CARREIRAS et al., 2002; BERNARDES et al., 2012) à dimensão local (EIPHANIO et al., 1994; KAWAKUBO et al., 2013).

Atualmente, com a disponibilidade de imagens de alta resolução espacial em níveis de detalhamento de poucos metros ou submétricos (por exemplo, Ikonos-2, QuickBird, GeoEye, WorldView-2, etc.), é possível extrair com bastante precisão informações do uso da terra antes negligenciadas pelas escalas mais generalizadas, já que esses dados orbitais de resolução submétrica são comparáveis aos de uma fotografia aérea (OLIVEIRA et al., 2009). Todavia, estas imagens ainda são pouco utilizadas, pois esses satélites de alta resolução não coletam imagens repetidamente de toda sua área de cobertura, mas adquirem suas imagens seletivamente, dependendo da solicitação do cliente (MACHADO; QUINTANILHA, 2008). Sendo que essa restrição se dá, principalmente, pelo elevado custo de aquisição.

Portanto, o aprimoramento de métodos destinados à extração de informações utilizando produtos com um nível de detalhamento no terreno entre 10-30 metros (sistemas Landsat-TM, ETM+; SPOT-HRV, HRVIR, HRG; CBERS-CCD; IRS-P6/LISS III etc.) é crucial para suprir a demanda de mapeamento ainda existente e que não foi suprida com as imagens de alta resolução. Tal questão é válida especialmente em aplicações de uso da terra e cobertura vegetal.

2. OBJETIVOS

Geral:

- Contribuir com os estudos de aplicação das técnicas de sensoriamento remoto para a classificação de uso da terra e de cobertura vegetal.

Específicos:

- Demonstrar o potencial e a limitação da técnica de classificação de imagens de satélite por meio de mineração de dados com o propósito de auxiliar na extração de informação temática relacionada ao uso da terra e cobertura vegetal.
- Verificar a possibilidade de se melhorar o desempenho da classificação, utilizando além das variáveis espectrais, atributos espaciais (como por exemplo, forma, tamanho, perímetro etc.) e topográficos (sub-bacias, altimetria, declividade, orientação das vertentes).
- Aplicar a metodologia de mineração de dados na sub-bacia do rio São Tomé como o propósito de colaborar com os estudos regionais de uso da terra e de cobertura vegetal no sul de Minas Gerais.

3. JUSTIFICATIVA

O presente trabalho faz parte de um projeto temático intitulado “Como o entorno interfere na conservação da biodiversidade de fragmentos florestais naturais ou antrópicos?”, que tem como objetivo conhecer o efeito de diferentes tipos de matrizes no entorno de fragmentos florestais naturais ou antrópicos, sobre a estrutura, diversidade e abundância de espécies da flora e fauna dos fragmentos. Um dos eixos do projeto temático consiste na descrição da paisagem com relação à distribuição, tamanho, formato e conectividade das matrizes.

Portanto, o aprimoramento de métodos voltados para a extração temática dos atributos da paisagem é de fundamental importância na pesquisa interdisciplinar na qual o presente trabalho se enquadra. A técnica de classificação por mineração de dados foi selecionada por dois motivos principais: 1) consiste em uma nova abordagem de classificação que ainda requer estudos para um melhor entendimento

(vantagens e desvantagens) da técnica e 2) permite inserir atributos que normalmente não são considerados no processo de decisão de uma classificação tradicional pixel a pixel.

A área de estudo, que corresponde à sub-bacia do rio São Tomé, foi selecionada por ser bastante representativa em termos de uso e cobertura vegetal no Sul de Minas Gerais. As propriedades são predominantemente de pequeno e médio porte, o que resulta em uma configuração da paisagem bastante fragmentada, o que a torna uma área interessante para a realização de um teste de classificação com uma nova concepção: a mineração de dados.

Variáveis espectrais, espaciais e atributos do terreno foram considerados na análise pelo fato de alguns cultivos presentes na área de estudo, especialmente o café, ser fortemente controlado pela disposição do relevo.

4. MATERIAIS E MÉTODOS

4.1. ÁREA DE ESTUDO

A área de estudada corresponde à sub-bacia do rio São Tomé, localizada no sul de Minas Gerais (Figura 1). O rio São Tomé constitui em um importante afluente do reservatório de Furnas, drenando uma área de aproximadamente 34 mil hectares que compreende parte dos Municípios de Machado, Serrania e Alfenas (SILVA et al., 2011). A nascente do rio São Tomé localiza-se no Município de Machado, a cerca de 1200 metros de altitude.

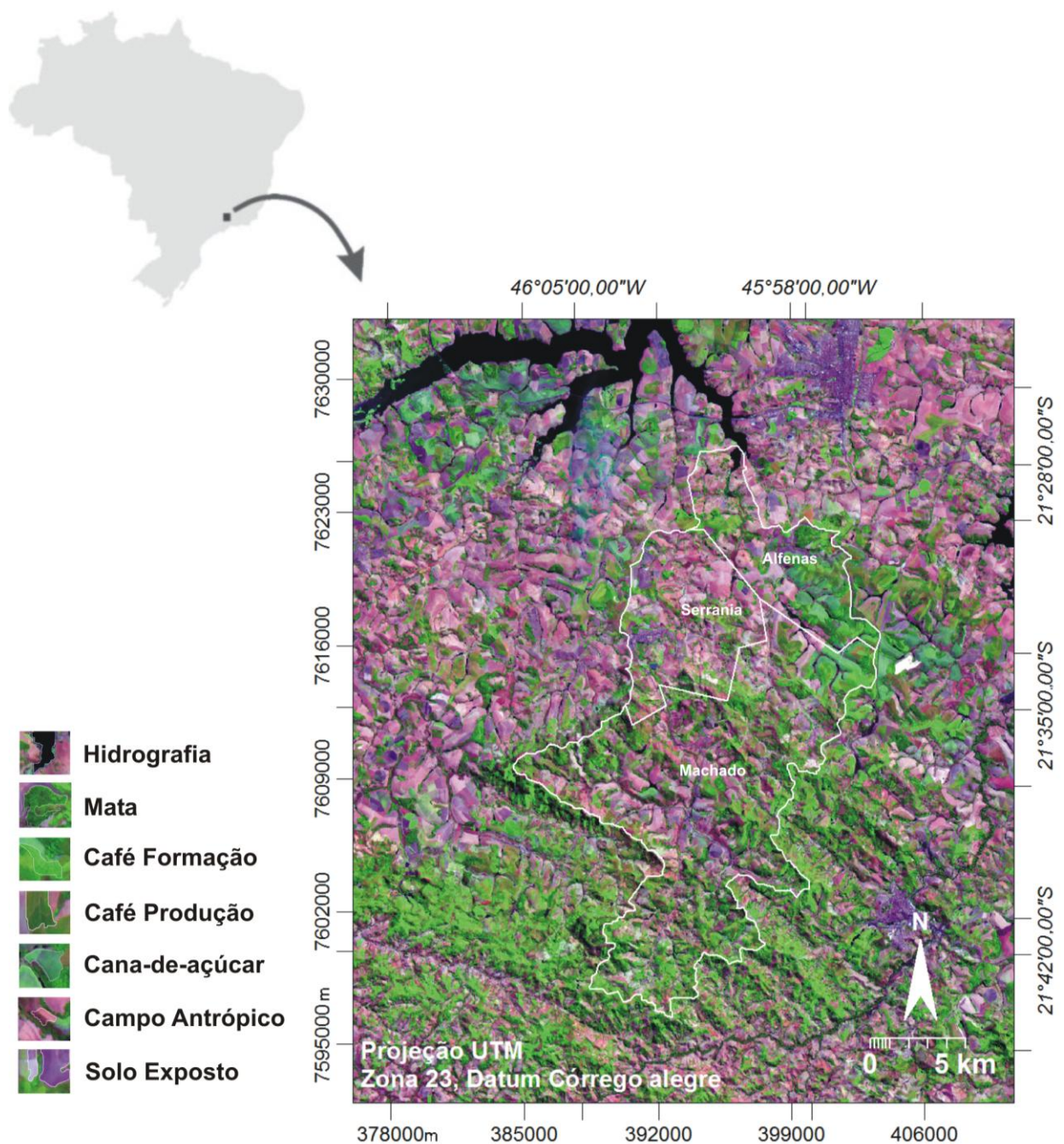


Figura 1. Mapa de localização da sub-bacia do rio São Tomé, sul de Minas Gerais. A imagem corresponde a uma composição colorida do sensor LISS III utilizando as bandas do infravermelho médio (b5), infravermelho próximo (b4) e vermelho (b3), combinadas em RGB.

A geomorfologia da região está incluída na unidade morfoclimática do domínio das regiões serranas de morros mamelonares, caracterizadas por morros arredondados em formato de “meia-laranja” (AB’ SABER, 1963).

Segundo o RADAMBRASIL (1983), a vegetação típica da região é formada por floresta estacional semidecidual ombrófila mista e cerrado. Já o clima é

subtropical, classificado como Cwa segundo Köppen, com verão úmido e inverno seco (PEEL et al., 2007).

Com relação ao histórico de ocupação, os municípios que compõem a região da sub-bacia do rio São Tomé tiveram um início de ocupação bastante parecido, já que, a princípio, eram apenas áreas de descanso para fazedores, tropeiros e boiadeiros. Somente nos séculos XVIII e XIX é que ocorre a fixação da população na região (IBGE, 2012).

O Brasil é o maior produtor de café do mundo (EMBRAPA, 2013) e Minas Gerais se destaca como o maior estado produtor, com 50% da produção nacional (RUFINO et al., 2010), colocando assim a região do sul de Minas como a maior produtora de café do país (SINDICAFÉ-MG, 2013). Nos três municípios pertencentes à sub-bacia contabilizou-se para o café um total de 26.597 ha de área plantada em 2010 e 27.057 ha em 2012. No campo da pecuária, o censo agropecuário (2006) contabilizou um total de 84.156 cabeças de gado para aquele ano; já em 2012 foram contabilizadas 78.525 (IBGE, 2013). Estes números colocam estas duas atividades como as mais importantes da região. Entretanto, mais recentemente, observa-se um avanço significativo do plantio da cana-de-açúcar, que alcançou em 2012, um total de 2.868 ha nos municípios da área de estudo (IBGE, 2013).

A Figura 2 abaixo demonstra quais foram os principais passos aplicados nas cartas topográficas e na imagem de satélite dentro da metodologia aplicada neste trabalho, para a confecção do mapa final de uso da terra e cobertura vegetal da área de estudo.

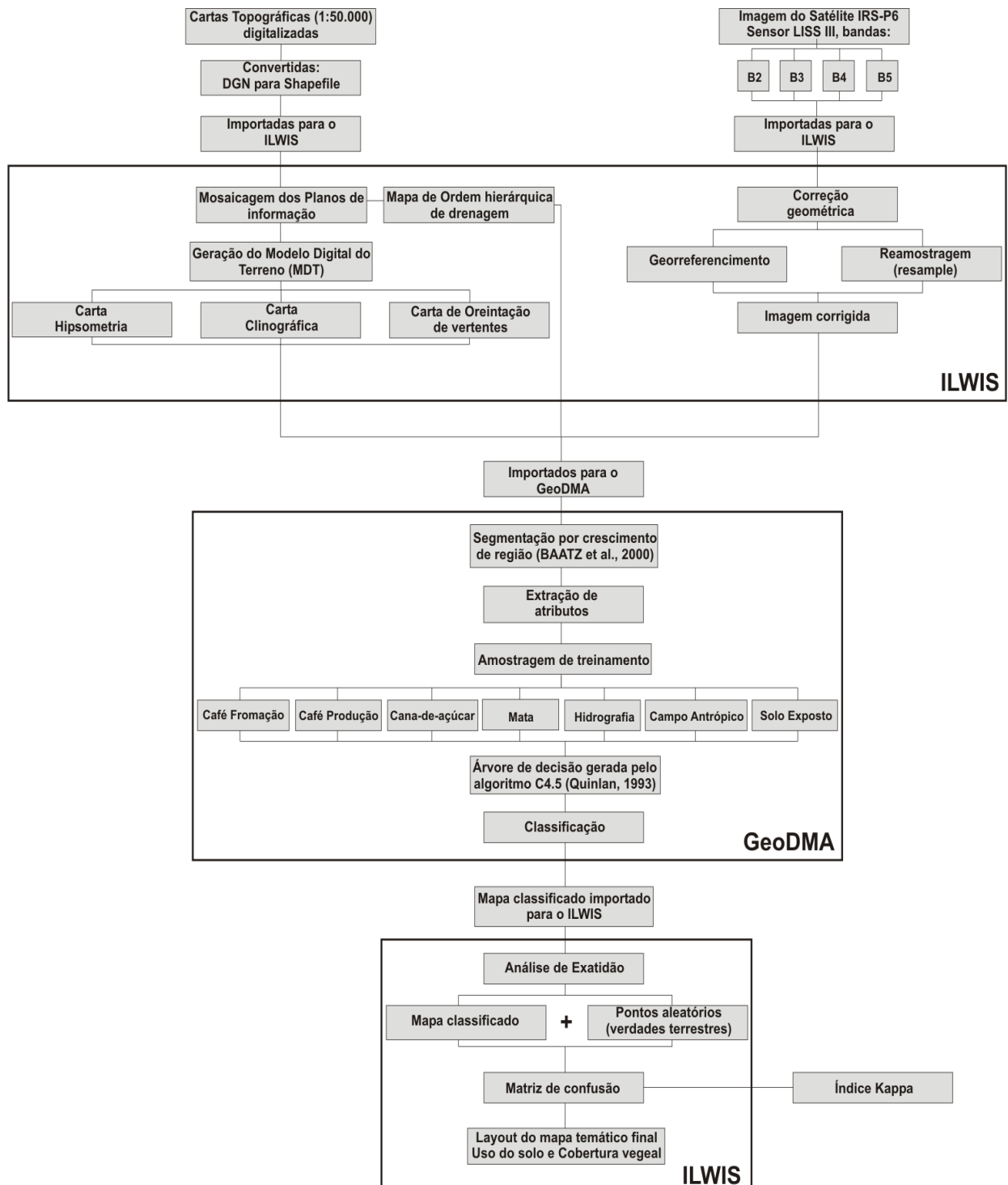


Figura 2. Fluxograma demonstrando os principais passos da metodologia aplicada neste trabalho.

4.2. ESTRUTURAÇÃO DA BASE CARTOGRÁFICA E CRIAÇÃO DE UM PROJETO EM SIG

Foram utilizadas como base cartográfica, quatro cartas topográficas na escala 1: 50 000 publicadas pelo IBGE, folhas: Areado (FOLHA SF-23-V-D-I-4), Alfenas (FOLHA SF-23-I-I-3), Campestre (FOLHA SF-23-V-D-IV-2) e Machado (FOLHA SF-23-I-III-1). Estas cartas, com os respectivos planos de informações (PI's) em formato DGN (*Design do MicroStation*), foram convertidas para o formato shapefile (.shp) e, em seguida, os PI's foram importados para o Sistema de Informação Geográfico ILWIS (*Integrated Land and Water Information System*).

Depois da importação dos PI's de curvas de nível, pontos cotados, drenagem e estradas, referentes a cada carta topográfica, os valores das coordenadas X,Y expressas em unidade de quilômetro, foram convertidas para a unidade métrica por meio de um algoritmo de conversão de coordenadas disponível no ILWIS. Em seguida, foi realizada a mosaicagem dos PI's com o objetivo de adquirir um produto contínuo que envolvesse toda a sub-bacia selecionada (Figura 3).

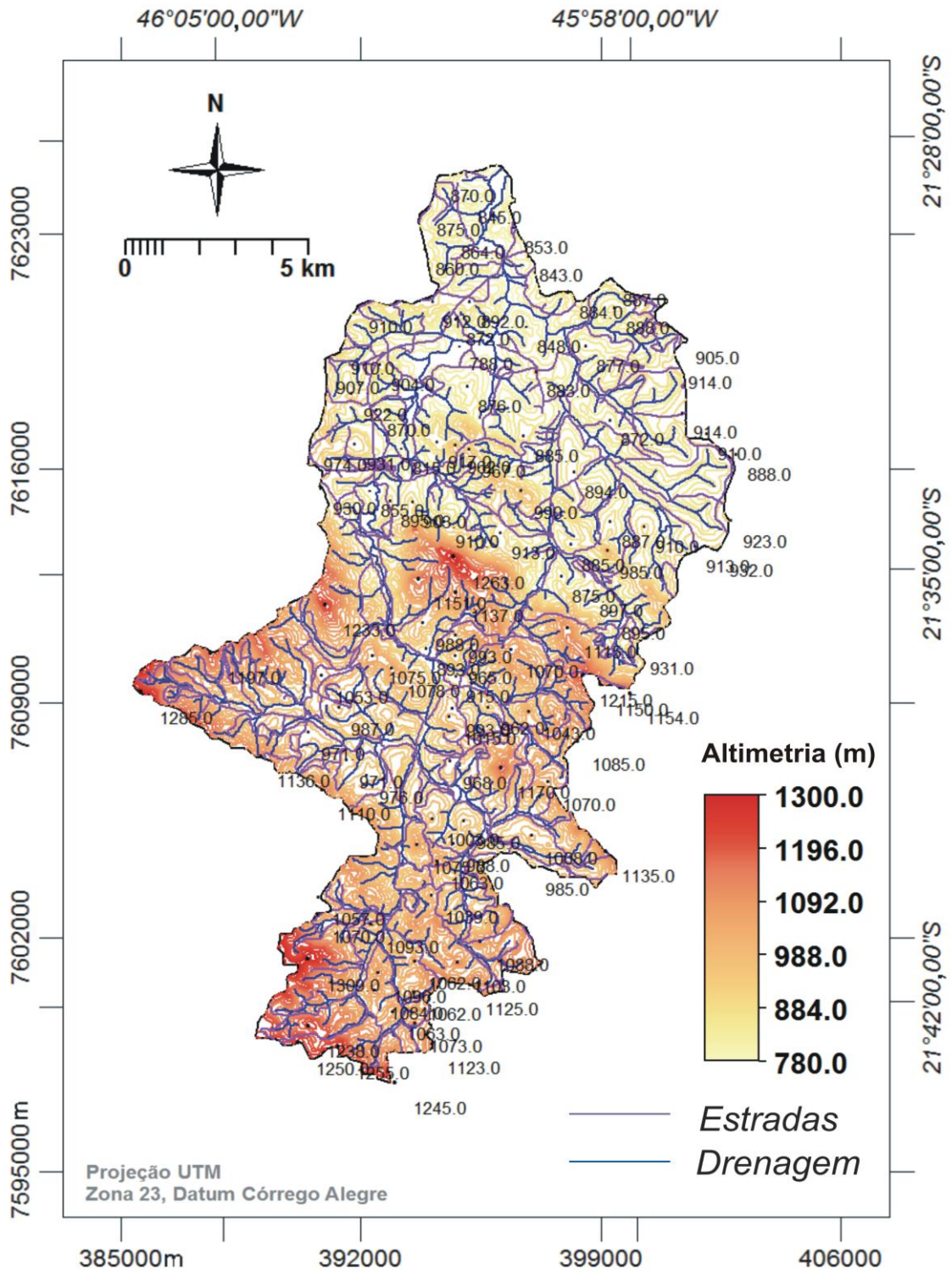


Figura 3. Área de delimitação da sub-bacia composta pelo mosaico dos PI's de curvas de nível, drenagem e estradas usados no SIG ILWIS.

A etapa seguinte consistiu na criação de um projeto no ILWIS, de um sistema de coordenadas, onde foi definido o retângulo envolvente da área de estudo e os parâmetros cartográficos (Tabela 1). Por fim, foi feita a edição das informações. Os PI's de drenagem e de estradas foram editados como tipo temático e as curvas de

nível e pontos cotados como tipo valor. Os erros de topologia também foram corrigidos nesta etapa por meio de análise visual, na tela do computador com o algoritmo *Check Segment* disponível no ILWIS.

Tabela 1. Parâmetros cartográficos definidos no projeto.

Retângulo Envolvente	Projeção	Datum Horizontal	Datum Vertical	Zona
Xmin.= 384306/ Xmax.= 404097	UTM	Córrego Alegre	Ibitura, SC.	23 S
Ymin.= 7596072/ Ymax.= 7627504				

4.3. DERIVAÇÕES DE INFORMAÇÕES DO TERRENO

Foram geradas informações do terreno com base nas informações das cartas topográficas implementadas no SIG ILWIS. Inicialmente foi gerado um modelo matemático contínuo do terreno e, em seguida, foram geradas também informações hipsométricas, clinográficas, de orientação das vertentes, de hierarquia de drenagem e de sub-bacias.

O Modelo Digital do Terreno (MDT) consiste numa representação matemática computacional da distribuição de um fenômeno espacial que ocorre dentro de uma região da superfície terrestre, sendo que dados de relevo são um dos fenômenos representados por esse modelo (FELGUEIRAS, 2001). As variações dadas pelo MDT e relacionadas à altitude do terreno (Figura 4) são descritas por três eixos: *x*, *y* que fornecem a posição espacial da superfície modelada (normalmente em unidades métricas como aquelas fornecidas pela projeção Universal Transversa de Mercator - UTM) e pelo *z* (atributo) que descreve a altitude do terreno.

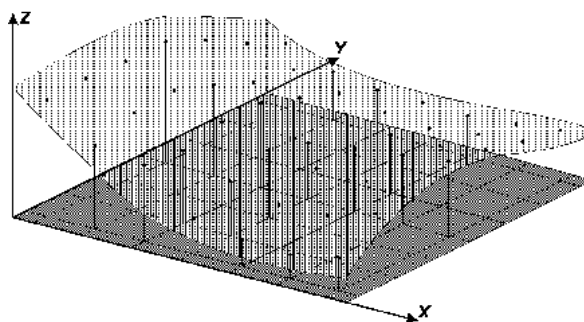


Figura 4. Modelo de superfície com os eixos *x*, *y* e *z* (DPI/INPE, 2013).

Estão descritos na literatura especializada diversos métodos de estimativas de superfícies contínuas que podem ser utilizados para modelar uma grande variedade de temas com diferentes níveis de precisão e de complexidade (BURROUGH, 1986; DESMET, 1997; BURROUGH; MCDONNELL, 1998; YAMAMOTO, 1998). A qualidade do modelo é influenciada por fatores relacionados, tanto à natureza dos dados, quanto aos processos de estimativa. A rugosidade do terreno, a densidade e a distribuição das amostras, além dos métodos de interpolação, são os principais fatores que exercem influência na qualidade dos modelos gerados (KUBIK; BOTMAN, 1976; LI, 1991; RAVIBABU; JAIN, 2008).

Para a geração do MDT, utilizando resolução espacial de 20 m, foram utilizadas informações de curvas de nível e pontos cotados com os devidos registros dos valores altimétricos. A geração do MDT foi feita utilizando um interpolador linear que se baseia na distância de Borgefors, que calcula o valor de menor distância em altura de cada contorno linear (Figura 5). Segundo Kawakubo et al. (2011), este algoritmo é dividido em duas etapas: a primeira, de conversão das linhas de contorno do formato vetorial para o raster. Após esta conversão, os pixels cobertos pela linha de contorno possuem valores de altitude e os demais pixels estão com valores indefinidos. Na segunda etapa é realizada a interpolação linear entre os pixels com valores de altitude, para obter as elevações dos pixels com valores indefinidos entre as linhas de contorno rasterizadas. As distâncias são calculadas, tanto para frente, quanto para trás, até que não ocorram mais mudanças (GORTE; KOOLHOVEN, 1990). Esse cálculo é feito utilizando a seguinte equação:

$$H_p = H_2 + (d_2 / (d_1 + d_2)) * (H_1 - H_2)$$

Onde, H_p é o valor da altura calculada para um pixel de saída; H_1 e H_2 são os valores da altura das linhas de contorno superior e inferior e d_1 e d_2 são as distâncias entre o pixel para as linhas de contorno superior e inferior.

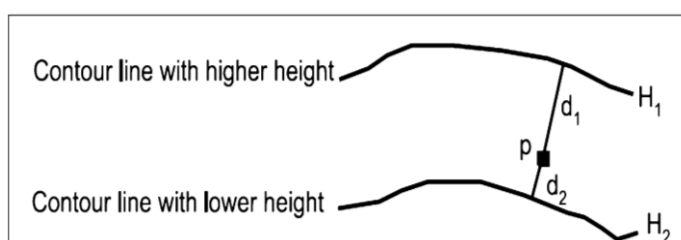


Figura 5. Cálculo da menor distância entre os contornos lineares H_1 e H_2 pelo método de Borgefors. Adaptado de Nijmeijer et al. (2001).

Por fim, os pontos cotados, também em formato raster são inseridos no MDT. A inserção destes pontos no modelo é importante para que, principalmente, os

topos de morro não fiquem achatados, criando uma aparência artificial da modelagem. A partir do MDT foram geradas, automaticamente, as cartas hipsométrica, clinográfica e de orientação das vertentes. Essas cartas foram criadas utilizando fórmulas preexistentes no ILWIS (Quadro 1).

Quadro 1. Fórmulas preexistentes no ILWIS e utilizadas para a elaboração das cartas Hipsométrica, Clinografia e de Orientação das vertentes, utilizando como apoio o guia de usuário do ILWIS (NIJMEIJER et al., 2001).

		Fórmulas	Resultado
		$\text{Combination} = \text{IFUNDEF}(\text{cotas}, \text{CurvasRasterized}, \text{cotas})$	Fórmula usada para gerar a combinação dos mapas de pontos e de segmentos já rasterizados para se calcular a hipsometria.
		$\text{dem1} = \text{MapInterpolContour}(\text{Combination})$	Fórmula usada para se confeccionar o mapa de altimetria de toda a área.
		$\text{Slope_percentage} = ((\text{HYP}(\text{Dx}, \text{Dy})) / \text{pixsize}(\text{dem_sub})) * 100$	Fórmula para gerar um mapa de porcentagem antes de confeccionar o mapa de declividade (Slope).
		$\text{Aspect} = \text{RADDEG}(\text{ATAN2}(\text{Dx}, \text{Dy}) + \text{pi})$	Fórmula usada para gerar um mapa de orientação de vertente de toda a área de entorno da bacia.
Cartas	Hipsométria	$\text{Dem_bacia} = \text{dem_sub} * \text{bacia_rec}$	Gerou o mapa de altimetria com os contornos da área de estudo com os valores de topo estáveis.
	Clinográfica	$\text{Slope} = \text{Slope_percentage} * \text{bacia_rec}$	Gerou o mapa de declividade com os contornos da área de estudo.
	Orientação de Vertentes	$\text{Aspect_bacia} = \text{Aspect} * \text{bacia_rec}$	Gerou o mapa de orientação de vertentes com os contornos da área de estudo.

As primeiras quatro fórmulas foram aplicadas em toda a área da imagem de satélite, onde HYP é a Hipotenusa, o DX é o filtro que detecta diferentes declives na direção X e o DY é o filtro que detecta diferentes declives na direção Y. Ambos os filtros foram aplicados no mapa Dem_sub (mapa de altimetria de toda a área da imagem) e depois utilizados na geração dos mapas clinográfico e de orientação das vertentes. Já RADDEG converte radianos para graus e ATAN2 retorna o ângulo em

radianos de dois valores de entrada; x é horizontal, y é vertical. Ambos são funções internas MapCalc / TabCalc. O valor de pi (π) é 3,141592653589....; as três últimas fórmulas do quadro foram utilizadas para gerar as cartas hipsométrica, clinográfica e de orientação das vertentes (NIJMEIJER et al., 2001), onde *bacia_rec* é o arquivo vetorial com os contornos da área da bacia.

As Figura 6, 7 e 8 ilustram os mapas confeccionados a partir das três últimas fórmulas do Quadro 1 e a figura 9 ilustra o mapa de hierarquia de drenagem e sub-bacias. Em (a), o mapa de hipsometria ilustra a variação dos valores de altimetria do terreno, da cota mais alta de até 1.309 m à cota mais baixa de 780 m. Em (b), o mapa clinográfico aponta a declividade da área de estudo em porcentagem, indo de 0 à acima de 45%. Já em (c), o mapa de orientação das vertentes aponta em qual direção dos pontos cardeais as vertentes da área de estudo estão direcionadas. O mapa (d) é o de hierarquia de drenagem e sub-bacias.

Para a elaboração da hierarquia de drenagem e sub-bacias, o procedimento foi realizado manualmente, utilizando como referência as informações das curvas de nível, da drenagem e da carta hipsométrica. Primeiramente, a rede de drenagem foi editada, utilizando a classificação de Strahler (1952), que define os cursos d'água em ordem hierárquica de primeira, segunda ordem e assim por diante. Neste trabalho, para a elaboração do mapa de hierarquia de sub-bacias, os limites destas foram digitalizados e, em seguida, poligonizados, utilizando um recurso de conversão de segmento para polígono no ILWIS, tendo a sub-bacia alcançado um nível hierárquico de até sexta ordem.

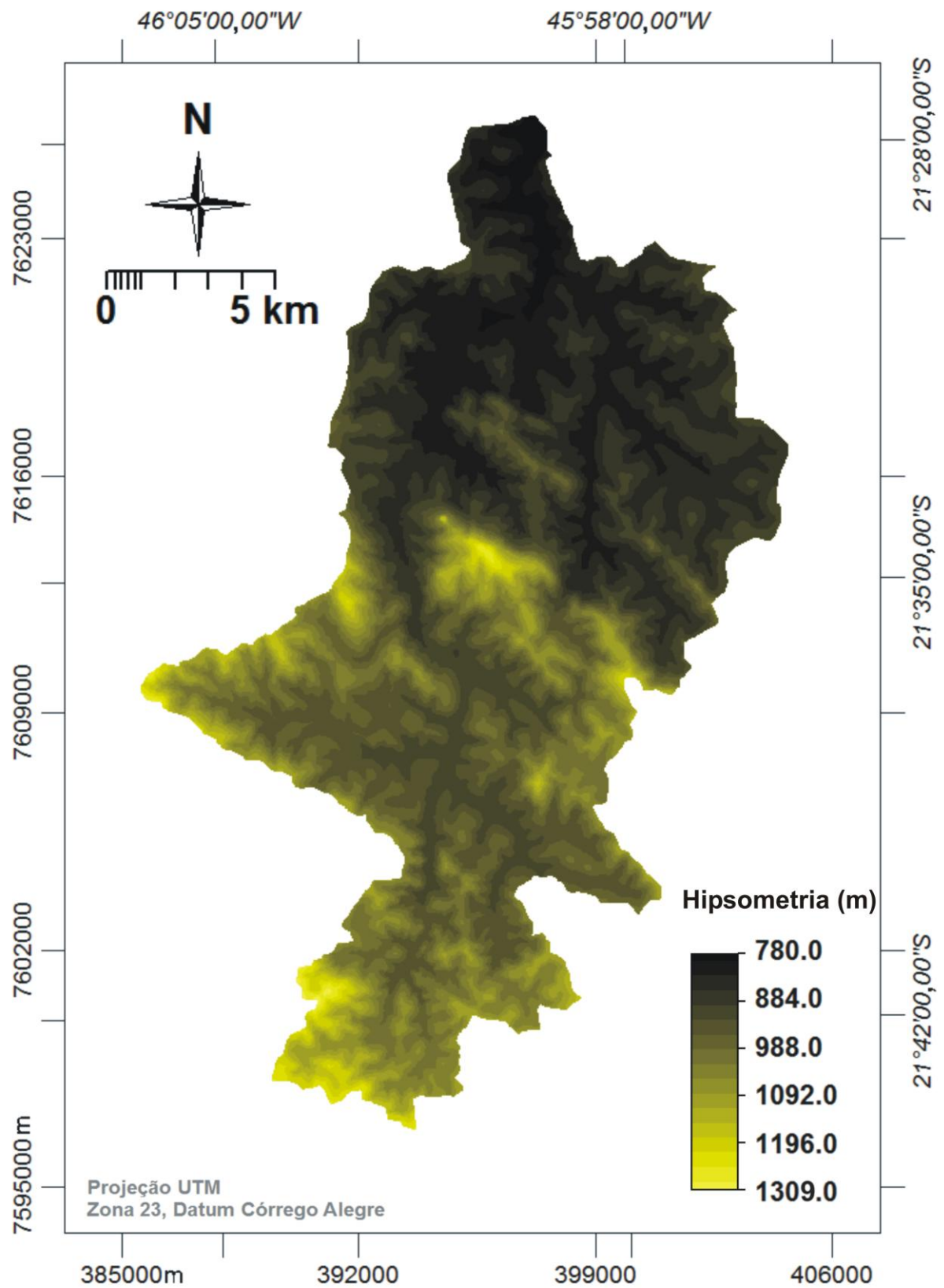


Figura 6. Mapa de hipsometria mostrando as diferenças de altitude do terreno, indo de 780m a 1309m.

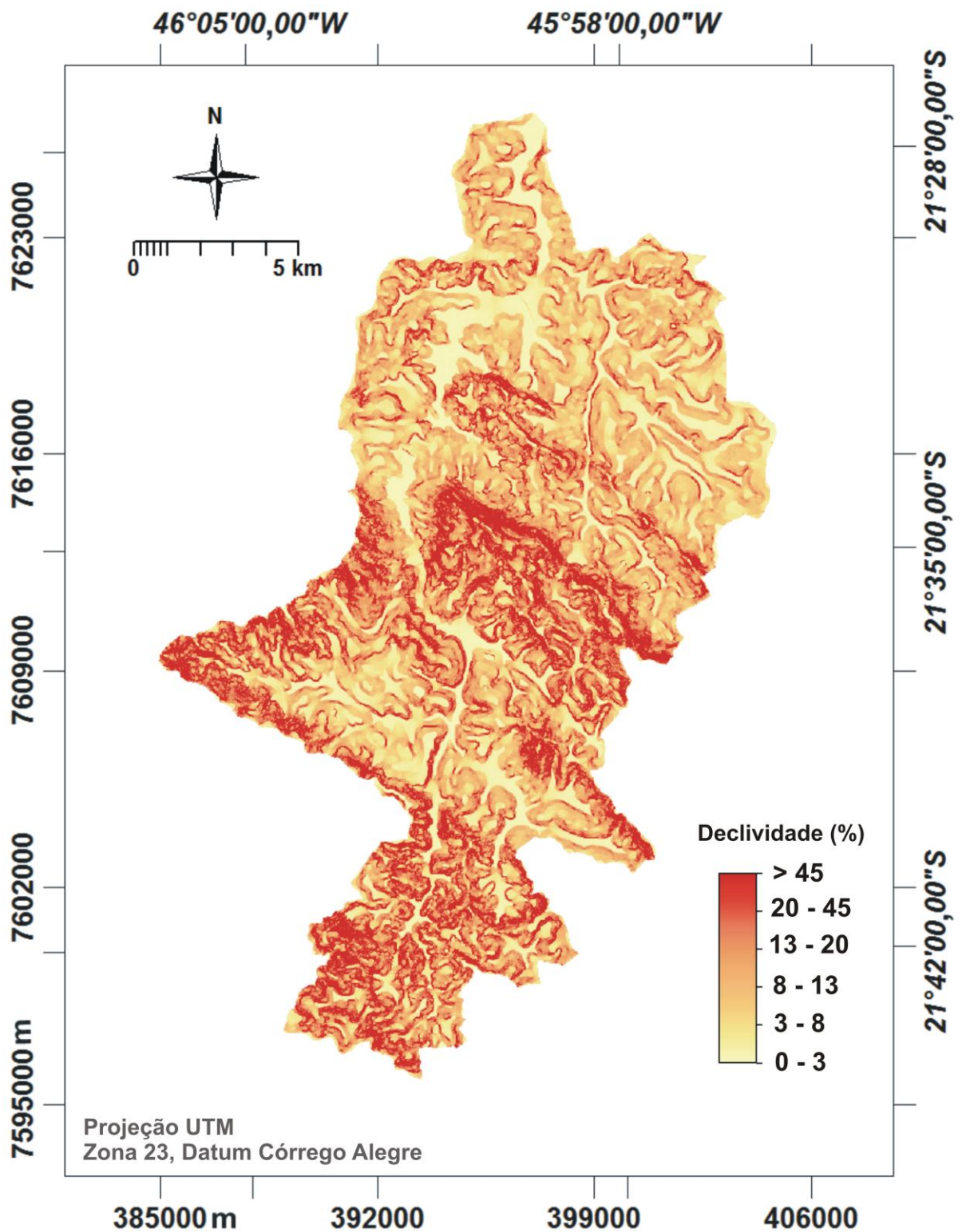


Figura 7. Mapa de declividade mostrando as diferentes inclinações do relevo dadas em porcentagem.

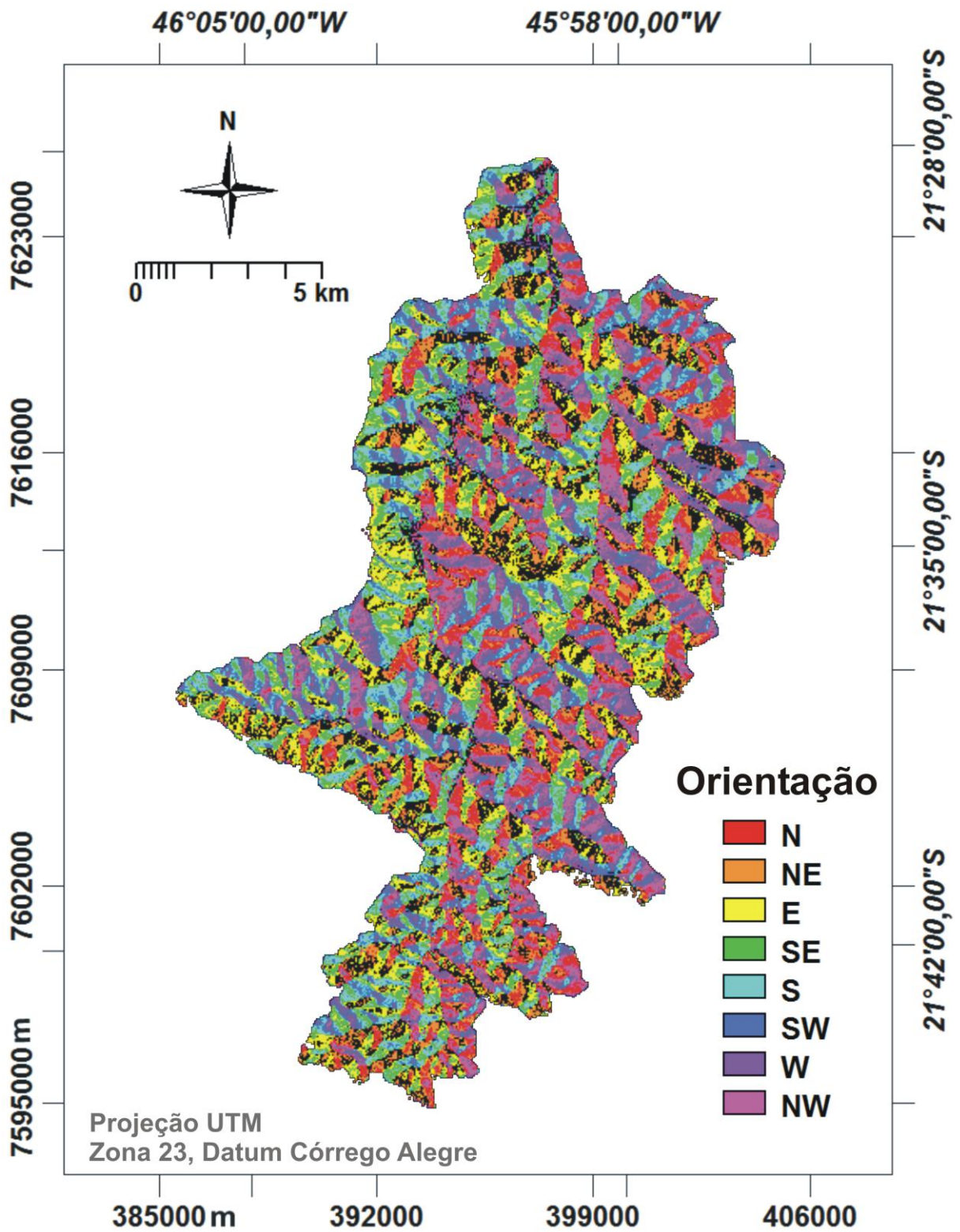


Figura 8. Mapas de orientação das vertentes apontando em qual direção as vertentes dentro da área de estudo estão direcionadas.

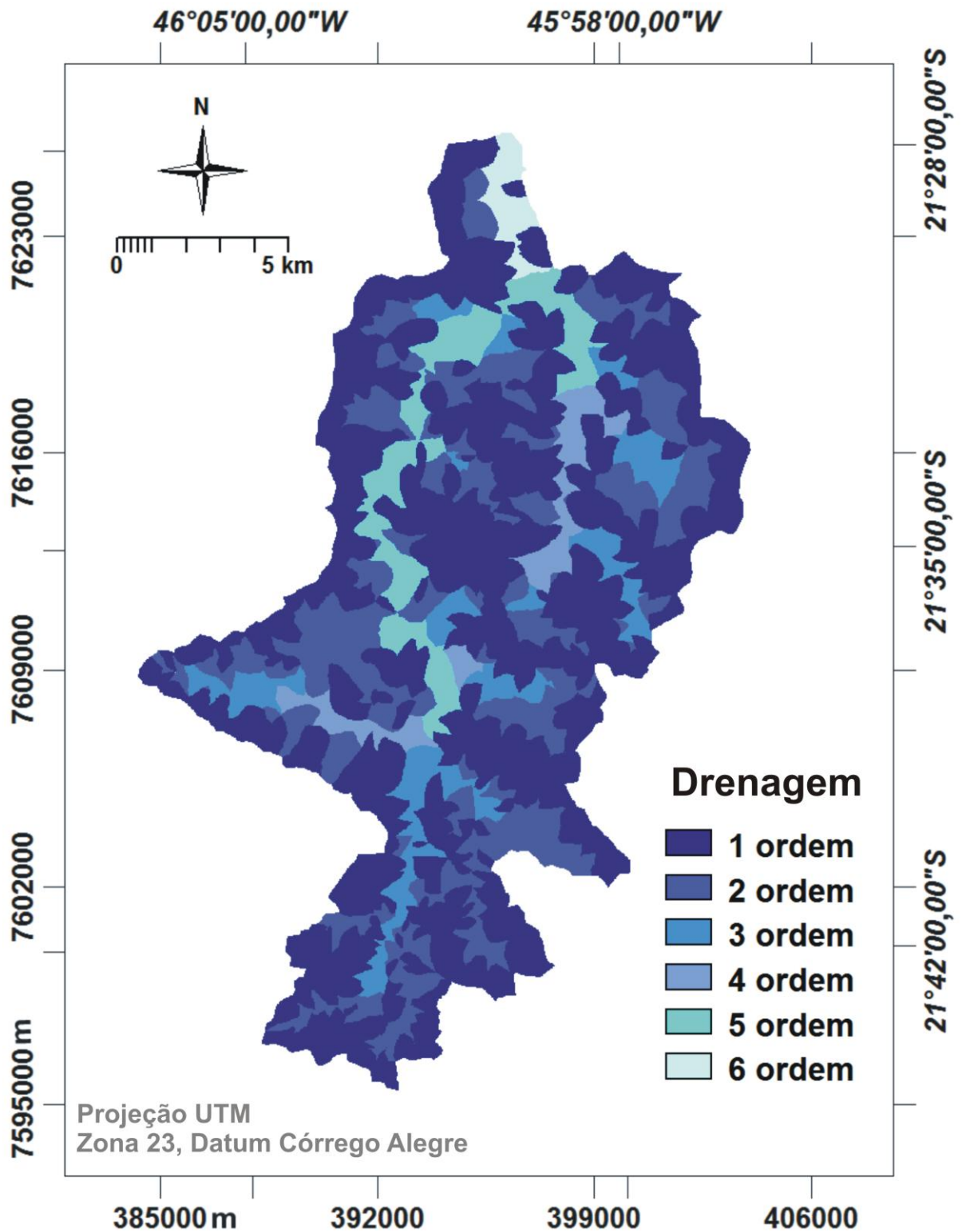


Figura 9. Mapa da ordem hierárquica dos cursos d'água na sub-bacia, mostrando que os cursos d'água atingiram a 6 ordem.

4.4. PROCESSAMENTO DAS IMAGENS

4.4.1. IMAGENS LISS III

A imagem obtida junto ao INPE e usada neste trabalho foi gerada pelo sensor LISS III (*Linear Imaging Self-Scanner*) que é um sensor do tipo *Pushbroom* e que possui quatro faixas espectrais (SILVA et al., 2013) que operam nos intervalos do verde (0.52-0.59 μm), vermelho (0.62-0.68 μm), infravermelho próximo (0.77-0.86 μm) e infravermelho médio (1.55-1.70 μm), com resolução espacial de 23,5 metros no terreno quantificado em 7 bits e com tempo de revisita de 24 dias (EMBRAPA, 2013). O sensor LISS III está a bordo do satélite IRS-P6 ou RESOURCESAT-1 que foi lançado em 2003 pela ISRO - *Indian Space Research Organisation* (Tabela 2).

Tabela 2. Características das bandas da imagem IRS-P6/ LISS III utilizada no trabalho.

Satel./ Sensor	Data	Az. Solar	Elev. Solar	Orb./ Ponto	Bandas	Range Espec. (μm)	Res. Esp. (m)	Res. Rad. (Bits)
IRS-P6/ LISS III	July, 31, 2010	35.5	42.4	331/93	B2	0.52 – 0.59	23.5	7
					B3	0.62 – 0.68	23.5	7
					B4	0.77 – 0.86	23.5	7
					B5	1.55 – 1.70	23.5	7

Fonte: DGI/ INPE, 2013.

As informações obtidas pelo sensor podem ser aplicadas em vários estudos, inclusive os de cunho ambiental como uso do solo, agricultura, vegetação dentre outros (INPE, 2013). As imagens LISS-III são compatíveis com outros sistemas de resolução moderada como as do Landsat 5 (*Thematic Mapper - TM*), 7 (*Enhanced Thematic Mapper Plus - ETM+*), 8 (*Operational Land Imager - OLI*) e CBERS (*High Resolution CCD Cameras*), que permite comparar as informações extraídas ao longo do tempo.

4.4.2. CORREÇÃO GEOMÉTRICA

A correção geométrica da imagem visa acertar as distorções geométricas causadas por vários motivos, principalmente aquelas resultantes da rotação da Terra e da instabilidade da plataforma (CRÓSTA, 2002). O procedimento de correção geométrica é um passo fundamental para a eliminação das distorções sistemáticas e

para permitir a integração das imagens com outras fontes cartográficas em SIG, inclusive possibilitar a comparação das imagens ao longo dos anos.

A forma mais comum de realizar a correção geométrica é por meio de georreferenciamento, que é feito identificando pontos de controle na imagem (ROQUE et al., 2006), utilizando como referência dados obtidos em um mapa, como no caso deste trabalho, que realizou o georreferenciamento utilizando como referência a base cartográfica estruturada ou por meio de um sistema de GNSS – *Global Navigation Satellite System* (SEGANTINE, 2001). Foram selecionados 15 pontos de controle distribuídos na imagem e que foram coletados nos cruzamentos de estradas, rios e feições geométricas visíveis na imagem.

Em seguida, foi aplicada uma relação de ajuste entre as coordenadas da imagem (expressas em linhas e colunas) com as coordenadas cartográficas, normalmente em unidades métricas e em projeção UTM. O ajustamento dos pontos foi feito utilizando a transformação afim, buscando no ajuste um erro inferior a 1 pixel. Essa transformação é um modelo matemático que consiste em corrigir uma imagem em relação à outra fonte de dados já corrigida, sendo que no caso deste trabalho foram as cartas topográficas. A imagem foi ajustada em um polinômio de 2ª ordem. A transformação afim é útil quando se deseja investigar deformações de escala ao longo de direções perpendiculares (D'ALGE, 2001).

Após o georreferenciamento, as imagens tiveram que ser reamostradas utilizando algum interpolador espacial. Os interpoladores espaciais têm como objetivo calcular o valor do nível de cinza da imagem corrigida. O ILWIS possuiu três interpoladores disponíveis: o vizinho mais próximo, o bilinear e o de convolução cúbica (Figura 10). O vizinho mais próximo tem a vantagem de não alterar o valor do pixel da imagem original, considerando no processo de interpolação o valor do pixel mais próximo. O interpolador bilinear e o de convolução cúbica consideram na interpolação os valores da vizinhança, abrangendo os quatro e dezesseis vizinhos, respectivamente. Apesar de alterar o valor da imagem original, estes dois interpoladores resultam em imagens com qualidade estética bem superior ao vizinho mais próximo.

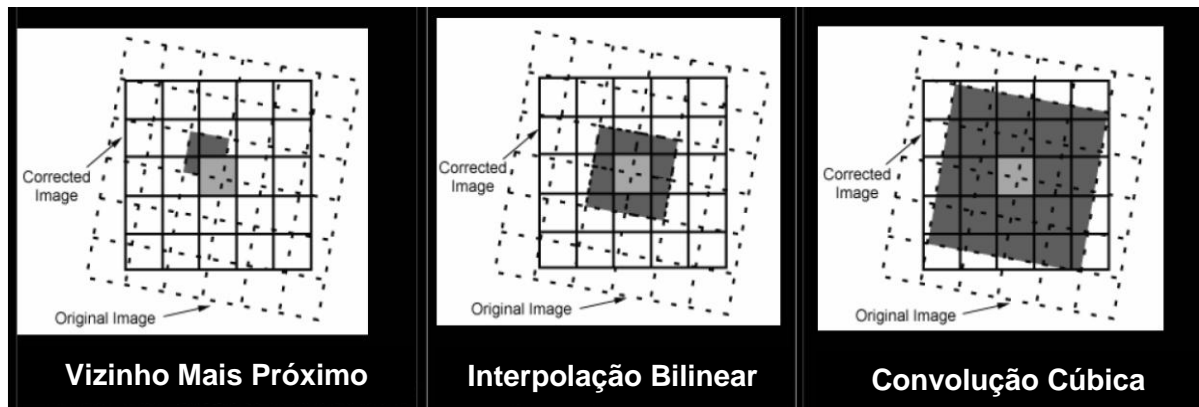


Figura 10. Os três interpoladores presentes no ILWIS, adaptado de Gonzalez & Woods (1992).

Neste trabalho foi selecionado o interpolador vizinho mais próximo para a correção geométrica das imagens pelo fato dos valores de brilho destas imagens não sofrerem alteração, característica importante na etapa de classificação. O espaçamento da grade definido no momento da reamostragem (*resample*) foi de 20 metros, resultando no final deste processo uma imagem com dimensão de 1.717 linhas e 1.454 colunas.

4.5. SEGMENTAÇÃO E CLASSIFICAÇÃO

4.5.1. SEGMENTAÇÃO

Segundo Bins et al. (1996), o processo de segmentação é uma das primeiras etapas na análise de imagens de SR onde a imagem é dividida em regiões que melhor representam objetos relevantes na cena. Na abordagem por segmentação, ao invés de se trabalhar a imagem no nível do pixel, se trabalha com o conceito de objetos que agregam pixels que possuem alguma propriedade em comum. São várias as vantagens de se trabalhar com objetos em comparação à abordagem tradicional pixel a pixel. Dentre os quais se pode citar a possibilidade dos segmentos guardarem informações espectrais adicionais como, por exemplo, os valores de média, moda, mediana, variância, etc., além de considerar atributos espaciais como a forma dos segmentos, perímetro, área, etc. (BLASCHKE, 2010), que são usados para treinar ou construir uma base de conhecimento para a classificação de imagens (KAWAKUBO et al., 2013).

Existem diferentes métodos de segmentação (HARALICK; SHAPIRO, 1985, PAL; PAL, 1993, KORTING et al., 2013). Por muitos anos, os procedimentos de segmentação têm sido foco de pesquisa nas áreas de análise de imagens (BAATZ et al., 2004). De acordo com Vand Der Sande et al. (2003), os algoritmos destinados à segmentação de imagens de SR podem ser grosseiramente divididos em três categorias: (1) detecção de bordas; (2) crescimento de regiões e (3) baseadas no conhecimento (KAWAKUBO, 2010).

A segmentação por detecção de bordas segmenta a imagem localizando as mudanças abruptas nos valores de intensidade dos níveis de cinza (PAL; PAL, 1993). Já a segmentação por crescimento de regiões (Figura 11) parte de um pixel “semente” que é comparado com o seus pixels vizinhos para determinar se são semelhantes, usando uma medida de similaridade. Se forem semelhantes, eles se agregam (BINS, 1996). A vantagem deste método é que nem todos os pixels pertencem a alguma destas regiões. Por fim, a segmentação baseada no conhecimento é a abordagem mais moderna existente para a extração de regiões em imagens de sensoriamento remoto. Neste tipo de segmentação, informações adicionais ou dados auxiliares, como por exemplo, mapas temáticos, podem ser incorporados no processo de segmentação para auxiliar na extração das regiões (KAWAKUBO, 2010).

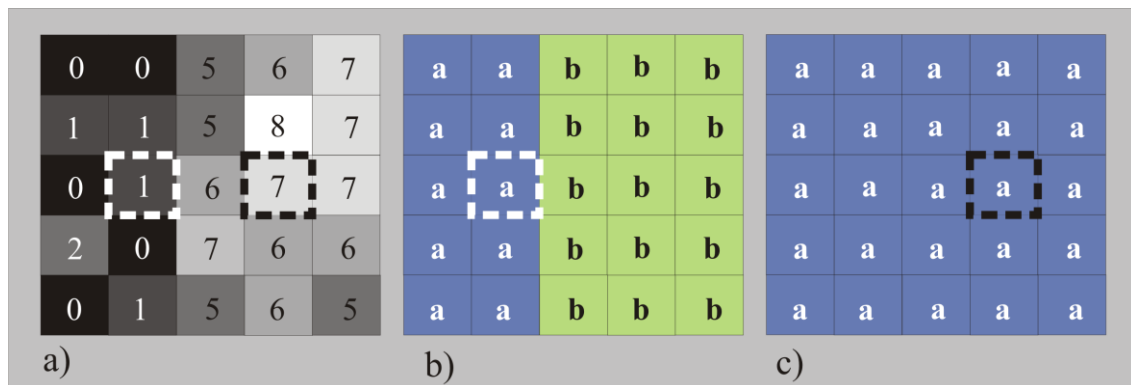


Figura 11. Exemplo de segmentação por crescimento de regiões (KAWAKUBO, 2010, adaptado de GONZALEZ E WOOD, 1992).

Na figura acima se tem em (a) Imagem original identificado os pixels sementes com valores 1 (coordenadas 3º linha, 2º coluna) e 7 (coordenadas 3º linha, 4º coluna); (b) resultado da segmentação rotulado como classes a e b usando um limiar T com diferença de menos 3 entre os níveis de cinza; e (c) resultado da

segmentação rotulado como classe adotando um T com diferença de menos 8. (KAWAKUBO, 2010, adaptado de GONZALEZ E WOOD, 1992).

Na segmentação foi utilizado o aplicativo GeoDMA (*Geographic Data Mining Analyst*), proposto por Körting et al. (2009; 2013). O GeoDMA integra métodos de análise de imagens com técnicas de mineração de dados (DPI/INPE, 2011). Segundo Korting et al. (2013) o GeoDMA funciona como um *plugin* do TerraView (Figura 12) e possui todas as fases de análise de imagens (segmentação, extração e seleção de atributos, mineração de dados e análise multitemporal).

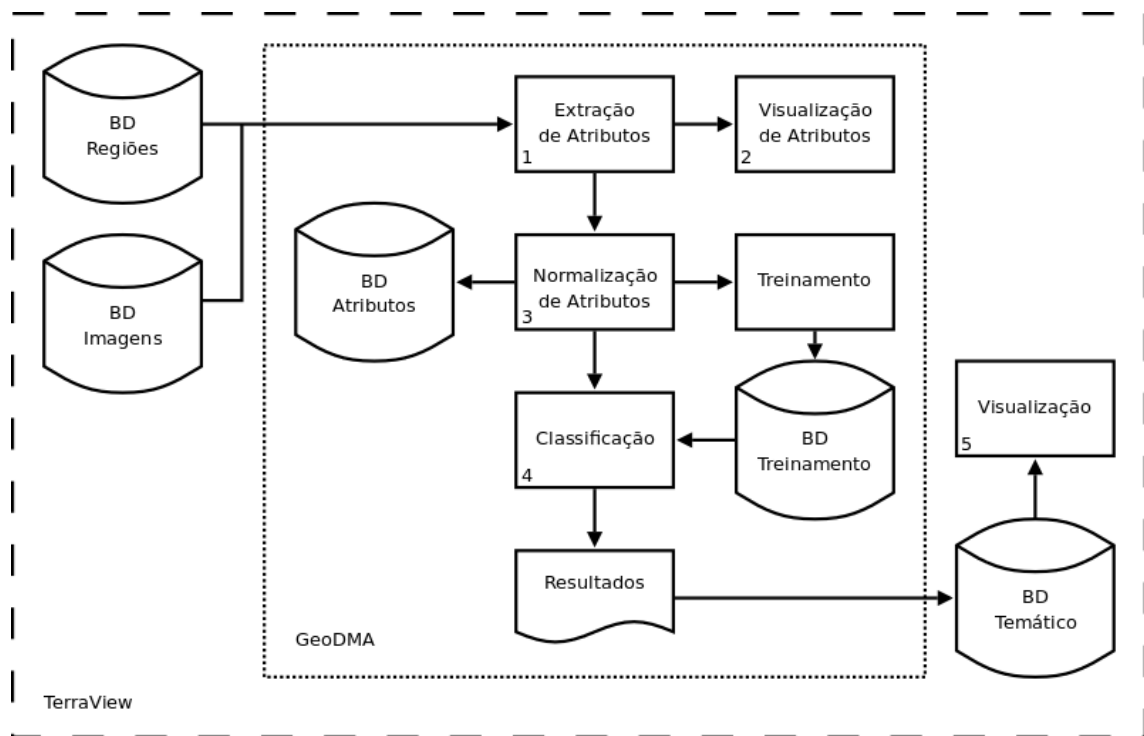


Figura 12. Interação do GeoDMA como *plugin* do TerraView, onde BD é o banco de dados trabalhado no programa. Adaptado de Korting et al. (2009).

Para realizar a segmentação, foi adotado o algoritmo por crescimento de região, baseado em Baatz et al. (2000), que incorpora o conceito de multirresolução. A multirresolução consiste em minimizar a heterogeneidade dos objetos de uma imagem (BAATZ et al., 2004; TIAN; CHEN, 2007), incorporando objetos menores dentro de objetos maiores com o objetivo de minimizar a heterogeneidade espacial (CHUBEY et al., 2006).

No GeoDMA, para gerar a segmentação, três critérios foram definidos: a compacidade, que trabalha com a densidade dos objetos compactados e não compactados dentro da imagem; a escala, que influencia no tamanho médio dos

objetos; e a cor, que tem relação com os valores espectrais das bandas da imagem (BAATZ, et al., 2004). Cada um deles tem influência no processo de segmentação, sendo que esses parâmetros recebem valores aleatórios definidos pelo intérprete para a geração da segmentação.

Foram realizados vários testes de segmentação com o objetivo de adquirir uma adequada ao conteúdo da imagem. Ou seja, buscou-se um resultado de segmentação que não fosse nem muito densa (*over-segmentation*), pois subdivide a imagem em regiões menores criando uma divisão excessiva da paisagem; e nem com falta de segmentos (*under-segmentation*), visto que resulta em uma classificação demasiadamente generalizada (BAATZ et al., 2004). Uma das consequências que pode ser observada de um excesso de segmentos é a perda de informações importantes relacionadas aos atributos espaciais do cultivo, como forma, área, perímetro etc. Por outro lado, quando ocorre uma falta de segmentos, classes distintas ou semelhantes podem ser agregadas e tratadas como sendo objetos comuns.

A análise dos resultados foi visual, verificando se os segmentos foram bem delimitados e ajustados ao conteúdo da informação presente na imagem (KAWAKUBO et al., 2013). Após a segmentação, foi iniciada a etapa de extração dos atributos. Além das bandas multiespectrais do sensor LISS-III (bandas do verde (b2), vermelho (b3), infravermelho próximo (b4) e infravermelho médio (b5)), foram também incluídas as informações derivadas do MDT, como a hipsometria, declividade, orientação das vertentes e ordem de sub-bacias hidrográficas. As informações topográficas foram incluídas porque o relevo exerce um forte controle no ordenamento territorial da bacia.

Estudos realizados por do Prado et al. (2013), na região de Muzambinho e Cabo Verde, nas proximidades de Alfenas-MG, também mostraram uma forte influência do relevo na distribuição espacial do plantio do café. Os autores observaram que o plantio do café distribuiu-se entre 800-1250 metros de altitude com predomínio na cota de 1000 metros. A declividade também exerce grande influência, sendo o café plantado normalmente da meia vertente em direção aos topos, evitando-se os fundos de vale em função dos maiores riscos de geadas.

Além das propriedades espectrais e topográficas, também foram extraídas propriedades texturais e espaciais de cada objeto delimitado pela segmentação. As propriedades texturais são importantes porque incorporam os conceitos de

variabilidade interna das classes. Já as propriedades espaciais são importantes porque as formas dos segmentos fornecem fortes indícios dos tipos de uso e de cobertura vegetal. As formas regulares, por exemplo, correspondem a diferentes plantios e as formas irregulares, as coberturas vegetais naturais.

4.5.2. CLASSIFICAÇÃO POR MINERAÇÃO DE DADOS

O objetivo das técnicas de classificação consiste em associar os pixels ou segmentos presentes na imagem a uma determinada classe de uso da terra ou cobertura vegetal. Existem diversos algoritmos destinados à classificação de imagens com variados níveis de complexidade e eficiência na extração de informação (GONZALEZ; WOOD, 1992; CROSTA, 2002; JENSEN, 2009).

Basicamente, podem-se agrupar as técnicas de classificação em: supervisionada, não-supervisionada e híbrida. Na classificação supervisionada o analista seleciona amostras de treinamento que são representativas das classes a serem mapeadas. Por meio de parâmetros estatísticos presentes no *software*, as amostras selecionadas são utilizadas como padrão de comparação aos demais pixels ou segmentos da imagem para decidir a qual classe eles pertencem. Na não-supervisionada o analista não precisa fornecer amostras de treinamento ao sistema, bastando definir o número de classes de saída (*clusters*). O agrupamento (ou *clustering*) é feito automaticamente pelo sistema, identificando-se as nuvens de pixels ou a dos valores dos segmentos que apresentam respostas espectrais semelhantes no espaço de atributos (CRÓSTA, 2002). Por fim, a classificação híbrida consiste em realizar uma não-supervisionada, como base para a seleção de amostras de treinamento, seguida de uma supervisionada (FLORENZANO, 2011).

O aplicativo GeoDMA utiliza-se da classificação supervisionada, ou seja, para realizar a classificação é necessária a coleta de amostras de treinamento com o objetivo de fornecer ao sistema parâmetros estatísticos de referência. Foram selecionadas amostras de sete classes, comumente encontradas na paisagem: corpo d'água, café formação, café produção, cana-de-açúcar, mata, pastagem e solo exposto. O café formação corresponde aos plantios de café em processo de formação, ou seja, em sua fase mais jovem, enquanto o café produção inclui os plantios em estágio de produção de grãos ou os plantios que foram podados.

Uma das abordagens mais promissoras é a classificação de imagens orientada a objeto, que trabalha com pixels agrupado em regiões, como já citado. Esta

classificação, baseada no conhecimento, simula o processo de fotointerpretação. Assim, critérios espectrais, texturais e de contexto são considerados no processo de decisão. De acordo com Vieira et al. (2012), uma etapa importante desta abordagem é a construção do modelo de interpretação de imagem (conhecimento). Neste modelo, o analista deve ter uma exata noção dos melhores atributos a serem utilizados na classificação. Condição esta, muitas vezes não atendida em virtude da dificuldade em gerenciar a grande quantidade de informações disponíveis e do nível de detalhamento exigido.

Como forma de superar tais limitações, a adoção de técnicas de mineração de dados surge como uma alternativa viável para geração automática de uma estrutura de conhecimento (SILVA et al., 2008; VIEIRA et al., 2012), já que segundo Korting et al. (2009), a mineração de dados utiliza as técnicas de aprendizado por máquina, permitindo trabalhar com um grande número de informações, e extrair ou evidenciar padrões nestes dados. No GeoDMA, a classificação supervisionada utilizada é mais complexa que as classificações tradicionais, pixel a pixel, já que utiliza a técnica de classificação por mineração de dados.

A palavra mineração nos remete ao ato de extrair alguma coisa, como por exemplo, na exploração de minerais na natureza. Em SR esse termo também tem o mesmo sentido; entretanto, aqui o que se extrai são informações dos bancos de dados. Neste trabalho foram realizados vários testes com a mineração de dados. Após a coleta das amostras das classes escolhidas, foi iniciado o processo de mineração. Esse processo consiste nas análises feitas pelo intérprete a cada classificação realizada e a cada árvore de decisão gerada pelo GeoDMA.

Quando um ou mais atributos afetavam de modo negativo a classificação, os mesmos eram retirados da classificação seguinte. Essa exclusão consistiu em deixar de fora atributos que atrapalhavam o processo de classificação, tentando dessa maneira, alcançar um melhor resultado na classificação final.

Foi notada que a retirada de um atributo, ou mais atributos, modifica a estrutura da árvore de decisão gerada na classificação seguinte (Figura 13), pois, as classes acabam sendo classificadas em atributos diferentes e em bandas diferentes. Além disso, quase sempre há a inclusão de um novo atributo.

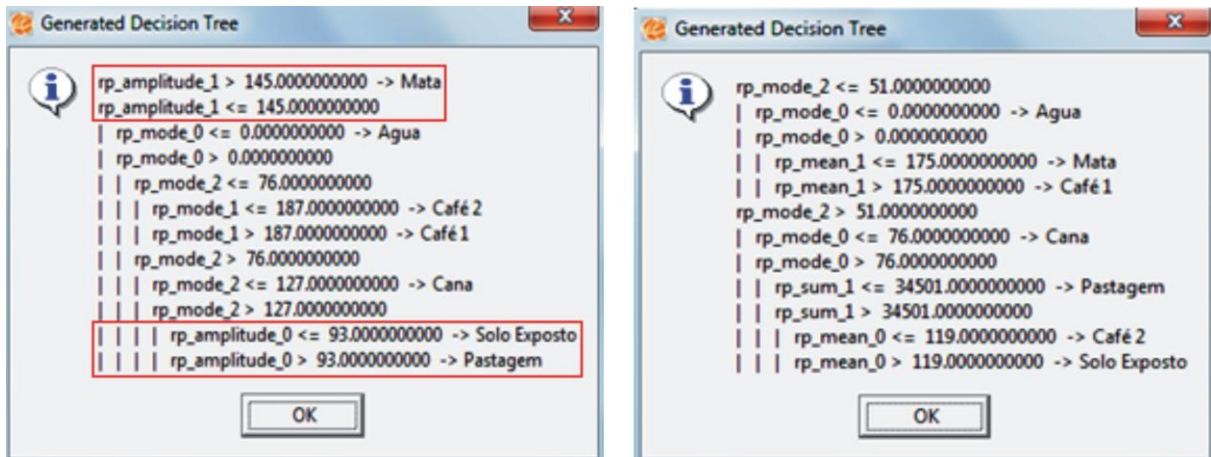


Figura 13. Na primeira classificação vê-se a presença do atributo amplitude. Esse atributo já não aparece na classificação seguinte, pois foi excluído do processo de mineração.

4.5.3. ALGORITMO C4.5

Para realizar o procedimento de classificação por mineração de dados, o GeoDMA utiliza um algoritmo para classificar as amostras selecionadas em relação às classes criadas pelo intérprete. O algoritmo utilizado pelo programa é o C4.5, desenvolvido por Quinlan (1993).

A principal vantagem de se usar o C4.5 é trabalhar com uma grande quantidade de dados e nenhuma hipótese de normalidade é adotada no processo de decisão. O resultado da classificação é expresso de maneira binária por meio de uma árvore de decisão como no exemplo da Figura 14. A escolha dos atributos é baseada em uma métrica de entropia denominada de ganho de informação. Os atributos com maiores ganhos de informação são selecionados na construção da árvore de decisão. As informações são organizadas e distribuídas em nós, ramos descendentes e folhas associadas às classes (GAMA, 2002), onde o primeiro atributo no topo da árvore representa a raiz; as linhas representam os ramos que se ligam aos nós que representam algum outro atributo que, se menor ou igual ao valor dado, se desloca para a esquerda e, quando maior, se desloca para a direita. Por fim, as classes classificadas nas extremidades dos ramos representam as folhas.

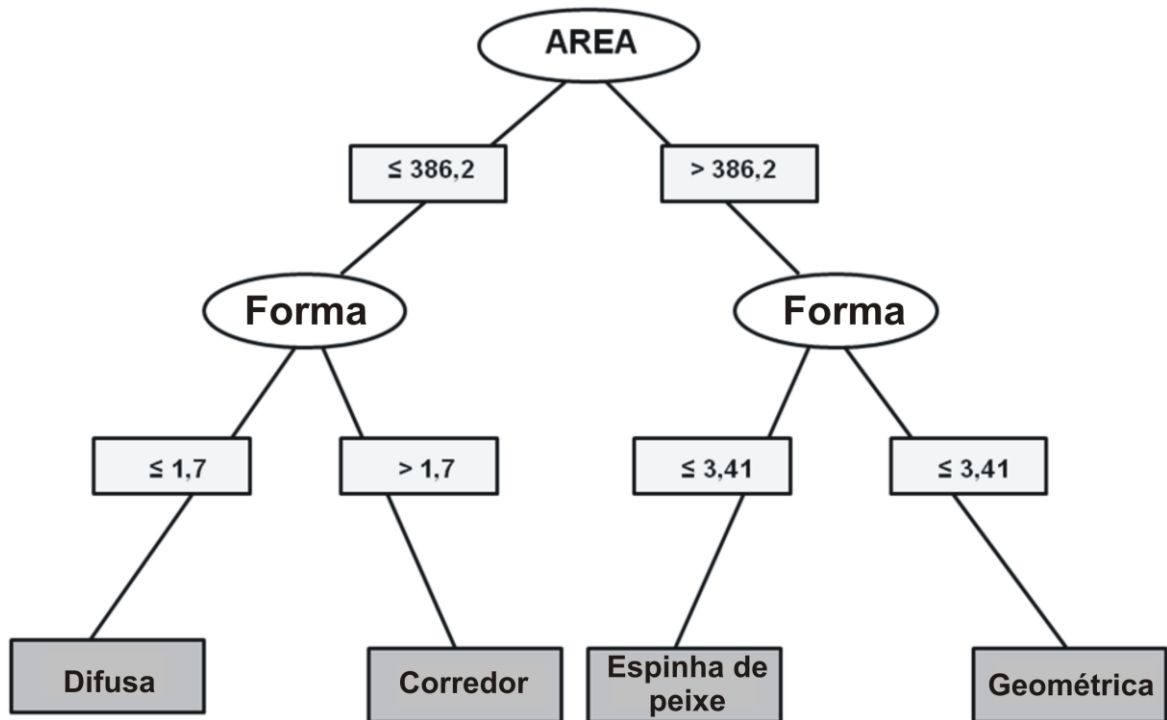


Figura 14. Exemplo da estrutura e aparência de uma árvore de decisão. Adaptado de Silva et al., 2008.

Os objetos da imagem são distribuídos e classificados por esse algoritmo, de acordo com as classes criadas pelo intérprete. A distribuição e classificação são feitas de modo automático pelo programa, baseando-se nos atributos espectrais e espaciais presentes no GeoDMA e selecionados pelo usuário. Além de não requerer suposições com relação à distribuição das amostras, a árvore de decisão é de fácil interpretação, possibilita considerar variáveis categóricas e numéricas e permite incorporar relações não lineares entre classes (FRIEDL; BRODLEY, 1997).

4.5.4. ANÁLISE DE EXATIDÃO

Foi realizado neste trabalho também uma análise de exatidão. Vale ressaltar aqui que exatidão se difere de precisão (Figura 15). Cabral (2004) descreve que exatidão é a maior ou a menor aproximação entre o resultado obtido e o valor real. Já a precisão está associada à dispersão dos valores resultantes da repetição das medições. O mesmo autor ainda compara exatidão e precisão como sendo um projétil disparado contra um alvo, onde a exatidão é o acerto próximo ou no centro do alvo, enquanto a precisão seria os disparos acertarem em pontos muito próximos entre si. Veiga et al. (2012) também explica essa diferença ao citar como exemplo um jogador de futebol que, cobrando dez pênaltis, acerta todos na trave direita,

sendo assim, o jogador foi preciso, mas não exato, já que não conseguiu acertar o gol que era o “verdadeiro valor”.

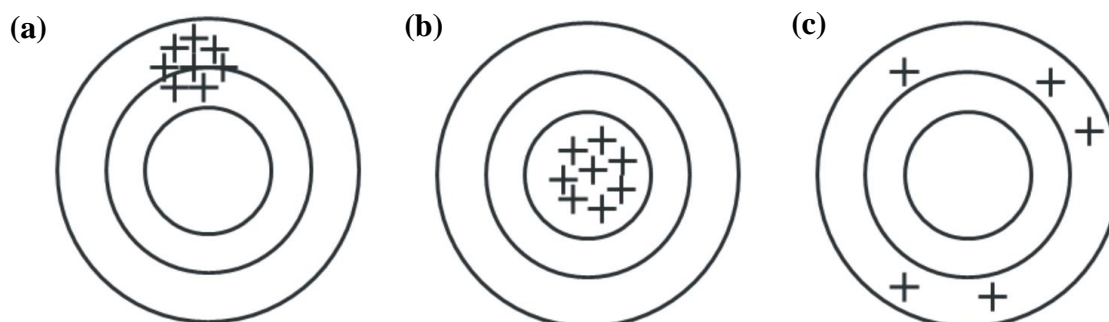


Figura 15. Precisão e exatidão. Em (a) preciso e inexato, (b) preciso e exato e (c) impreciso e inexato. Adaptado de Veiga et al. (2012).

A análise de exatidão da classificação é uma etapa crucial para verificar os acertos e os erros contidos no mapeamento. Esta análise é feita por meio de uma tabela denominada de matriz de confusão. Para a construção desta matriz, dados de referência, distribuídos aleatoriamente, são coletados. Em seguida, o resultado da classificação é combinado com os dados de referência por meio da técnica de tabulação cruzada.

A matriz de confusão é uma matriz quadrada formada por linhas e colunas onde são dispostos os resultados da classificação e as amostras de referência. Os valores distribuídos na diagonal principal desta matriz correspondem aos acertos obtidos no mapeamento. Os valores fora da diagonal principal correspondem aos erros que podem estar associados à inclusão ou exclusão de classes (CONGALTON, 1991). Os erros de inclusão e exclusão são os referentes às classes dentro da matriz de confusão. O erro de inclusão se refere a objetos presentes ao longo da linha e que foram incluídos erroneamente em outras classes, enquanto o erro de exclusão se refere a objetos presentes ao longo da coluna e que foram incluídos erroneamente em outras classes.

Utilizando a técnica de tabulação cruzada, os dados de referência foram combinados com o resultado da classificação. Foram extraídas porcentagens de acertos globais (proporção de pontos dispostos na diagonal principal da matriz) e individuais, bem como os respectivos erros associados com o objetivo de avaliar as causas das confusões. Também foi utilizado o índice kappa como referência de qualidade do mapeamento. De acordo com Congalton e Green (1993), o índice

kappa é recomendado como uma medida aceitável de precisão, sendo mais confiável que apenas a média da exatidão global, pois utiliza em um único índice todos os valores da matriz de confusão. Detalhes de como são calculados os índices mencionados são descritos com maior detalhe em Congalton (1991).

Para a construção desta matriz foram coletados três mil pontos de verdades terrestres (Figura 16) distribuídos aleatoriamente na área de estudo. Congalton (1991) afirma que pelo menos cinquenta pontos de referência são necessários para cada classe mapeada para que a análise de exatidão tenha significado estatístico. A checagem destes pontos foi feita em trabalhos de campo e com auxílio do Google Earth.

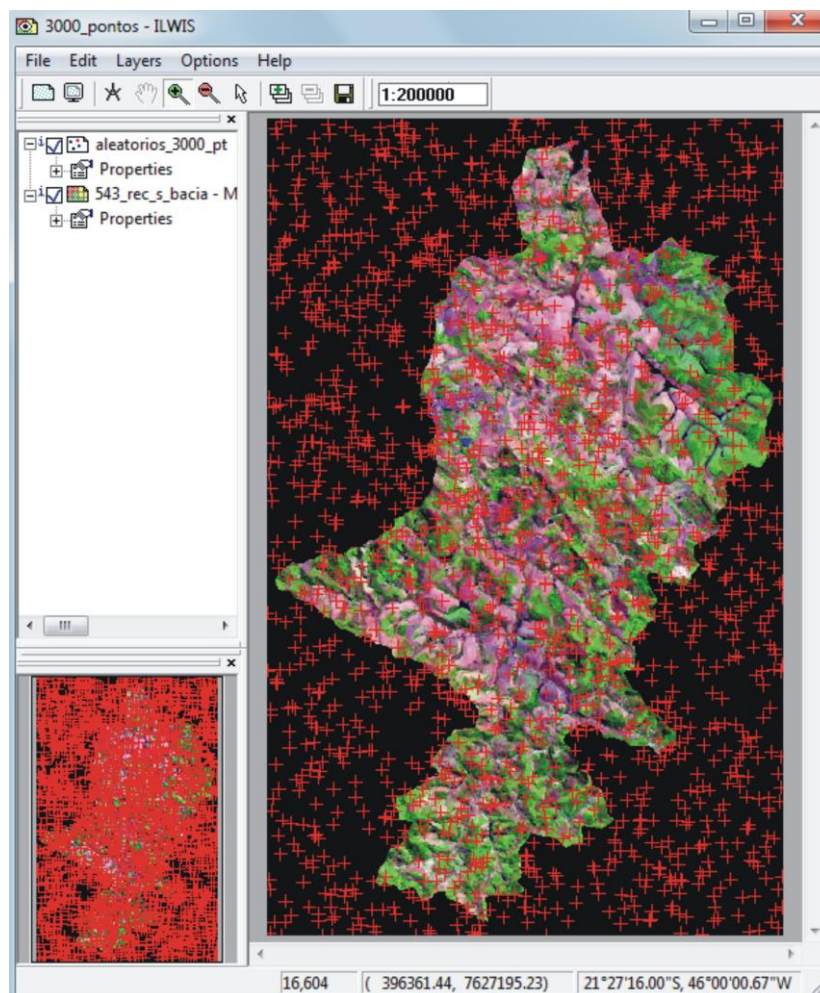


Figura 16. Pontos aleatórios de verdades terrestres a serem validados no ILWIS.

5. RESULTADOS

Após vários testes de segmentação, foram adotados como parâmetro de segmentação os seguintes valores: 40 (compacidade), 30 (escala) e 70 (cor). A Tabela 3 mostra os valores alcançados em uma segmentação com *over-segmentation*, em outra com *under-segmentation* e a segmentação ideal usada no trabalho. Já a Figura 17 ilustra alguns exemplos de resultados utilizando diferentes valores de entrada.

Tabela 3. Utilizando diferentes parâmetros de compacidade, escala e cor gerou-se em (a) uma segmentação alta com 13505 polígonos, em (b) uma baixa segmentação com 406 polígonos e em (c) melhor segmentação com 2499 polígonos.

Segmentação	Compacidade	Escala	Cor	Num. de Polígonos	Escolha
Alta	30	10	40	13505	x
Baixa	30	40	10	406	x
Melhor	40	30	70	2499	✓

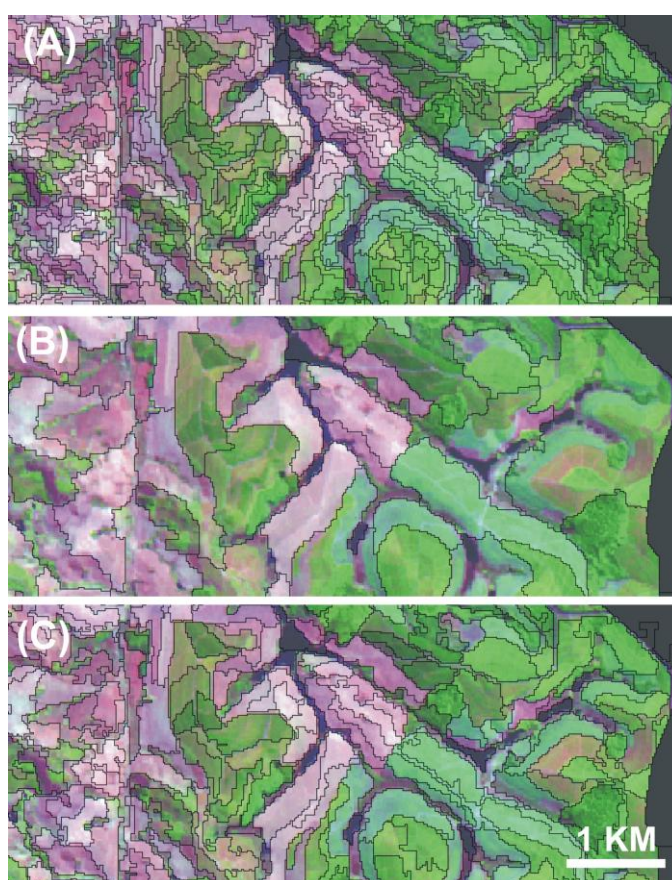


Figura 17. Resultados da segmentação utilizando diferentes parâmetros de compacidade/escala/ cor. Observa-se em (A) excesso de segmentos, em (B) falta de segmentos e em (C) um particionamento da paisagem mais equilibrado.

Conforme já mencionado, a definição dos polígonos da segmentação foi feita de maneira visual observando o particionamento da imagem. Como regra geral, os melhores valores são aqueles que conseguem delimitar com eficiência os objetos contidos na imagem sem que haja falta ou excesso de polígonos. Outra forma de avaliação da segmentação é por meio de estatística espacial realizada por Câmara et al. (2003) e Spindola et al. (2007) que utilizaram o índice Moran para definir quais são os melhores valores de segmentação utilizando como critério a dependência espacial.

Gerou-se na imagem segmentada uma classificação por mineração de dados, utilizando as amostras coletadas na fase de treinamento da imagem. Com relação a esta etapa de mineração de dados, a árvore que apresentou o melhor resultado é ilustrada na Figura 18. Conforme pode ser observado, o parâmetro dominante utilizado para classificar as regiões foi à moda das propriedades espectrais.

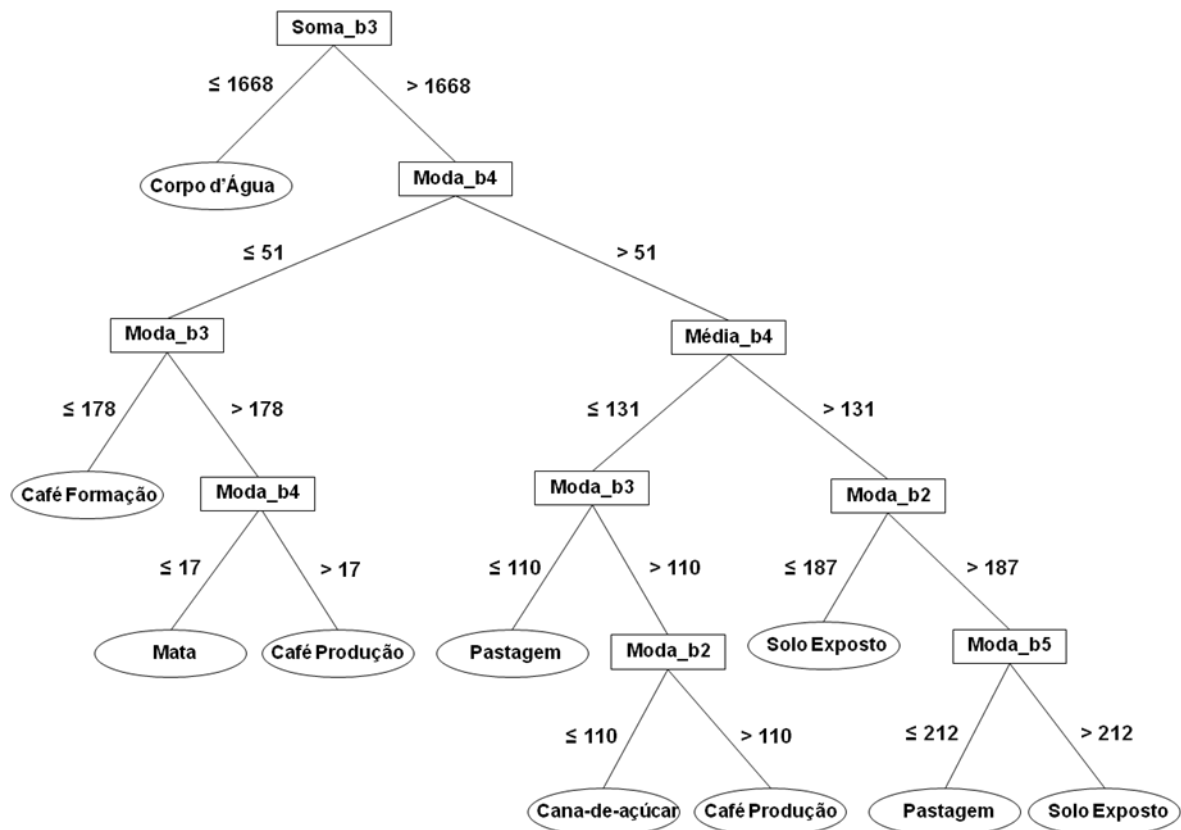


Figura 18. Árvore de decisão para a classificação por mineração de dados gerada no GeoDMA utilizando o algoritmo C4.5.

Analisando a árvore gerada, a primeira divisão separou a classe corpo d'água das demais utilizando a soma da banda 3 (comprimento de onda do vermelho). Os

objetos com valores iguais ou abaixo de 1668 foram classificados como corpo d'água e os objetos com valores acima foram subdivididos.

Na divisão seguinte, objetos com valores de moda na banda 4 (infravermelho próximo) iguais ou abaixo de 51 e com valores de moda na banda 3, iguais ou inferiores a 178, foram classificados como café formação. Os objetos com valores acima de 178 e com valores iguais ou inferiores a 17 na banda 4 foram classificados como mata. Os valores maiores que 17 foram então classificados como café produção.

Os objetos com valores acima de 51 na banda 4 foram subdivididos, utilizando a média de seus valores. Objetos com valores iguais ou inferiores a 131 e que apresentaram, ao mesmo tempo, valores iguais ou inferiores a 110 na moda da banda 3, foram rotulados como pastagem. Já os objetos com valores acima de 110 e com valores de moda da banda 2 (canal do verde) iguais ou inferiores a 110 foram classificados como cana-de-açúcar. Os valores acima de 110 foram novamente classificados como café produção.

Objetos com valores acima de 131 na média da banda 4 e com moda na banda 2, iguais ou inferiores a 187, foram classificados como solo exposto. Os objetos com valores acima de 187 e com valores iguais ou inferiores a 212 na moda da banda 5 (infravermelho médio) foram rotulados como pastagem e os valores acima de 212 como solo exposto.

Conforme pode ser observado, as classes pastagem, solo exposto e café produção foram classificadas em dois momentos distintos nas subdivisões da árvore de decisão. Isto aconteceu, provavelmente, em virtude da heterogeneidade destas classes. A classe pastagem inclui tanto a vegetação rasteira com baixa atividade fotossintética (pastos secos) quanto os pastos sujos e banhados. A classe solo exposto também apresenta grande diversidade espectral, pois inclui os solos secos e os solos preparados para o cultivo. Já a classe café produção engloba tanto o café em plena atividade produtiva (caracterizada por índices mais altos de área foliar e de sombra de dossel) quanto o café velho ou podado (caracterizado pela maior presença de galhos secos e menor abundância de folhas verdes e de sombra).

O resultado da classificação pode ser visto na Figura 19, que ilustra a distribuição do uso do solo e da cobertura vegetal. Observa-se no mapa que, ao longo de toda a área de estudo, o uso do solo é bastante particionado e a cobertura vegetal é bastante fragmentada e espalhada ao longo da sub-bacia. Dentro da

escala de análise utilizada neste trabalho (resolução espacial de 20m), nota-se que não só ao longo do rio São Tomé, mas também em todos os outros cursos d'água que compõem a sub-bacia uma grande ausência de mata ciliar, estando a mesma presente em poucos lugares. O café dividido em formação e produção é a principal cultura agrícola da bacia. Essa cultura ocupa principalmente as cotas mais altas do relevo, sendo cultivada em vertentes e topos de morro. O café produção visivelmente é o mais presente na área mapeada.

As outras classes que se fazem presentes de modo significativo na bacia são o solo exposto e a pastagem. Essas duas classes ocupam áreas às margens dos cursos d'água, tomando o lugar, como já citado, da mata ciliar. A pastagem é usada principalmente para a criação de bovinos, sendo esta outra atividade que se destaca na área da sub-bacia. O cultivo de cana-de-açúcar, apesar de ter mostrado crescimento na região, ainda tem impacto pouco significativo, ocupando de modo mais intenso uma área a nordeste do rio São Tomé.

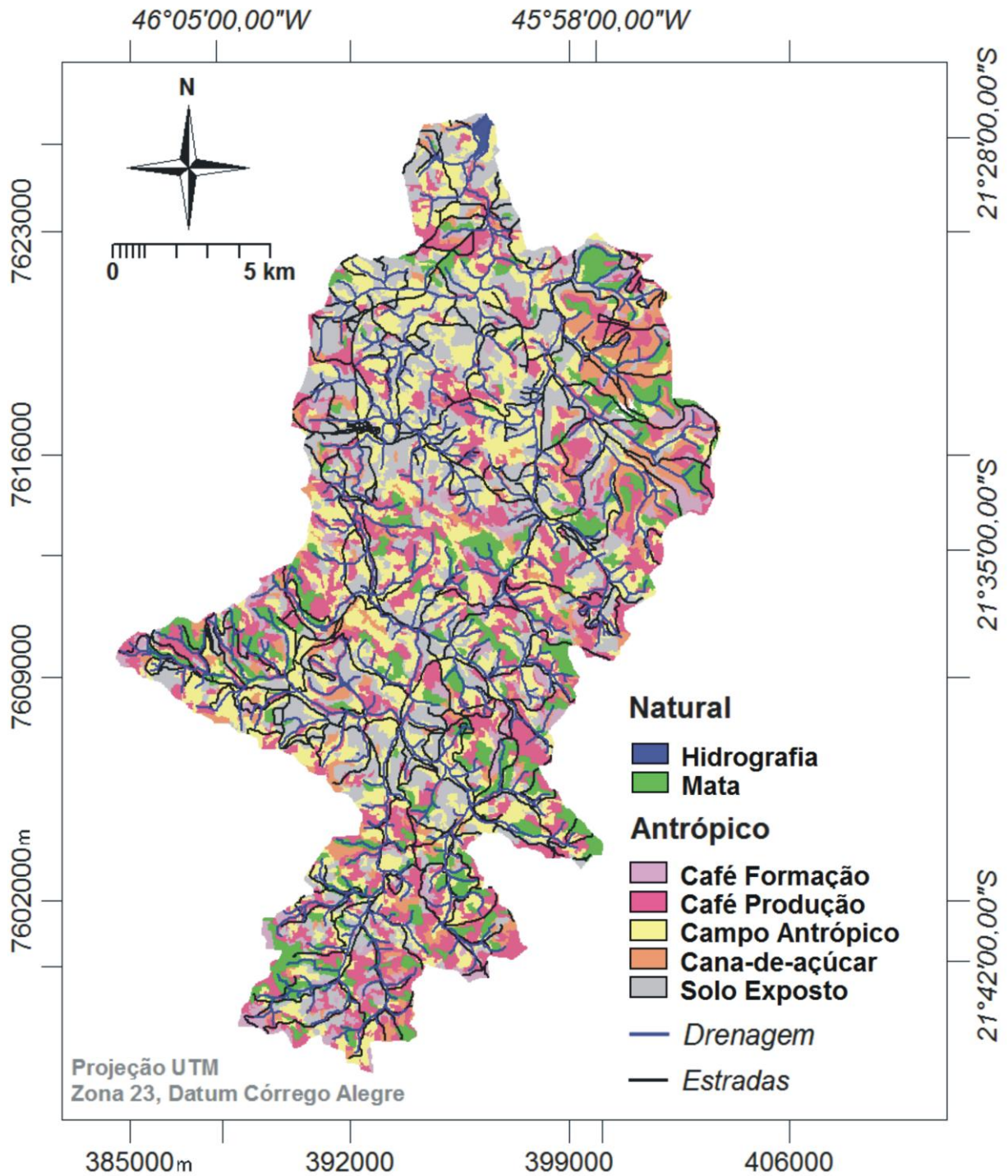


Figura 19. Mapa final de uso da terra e cobertura vegetal confeccionado para a área da sub-bacia hidrográfica do rio São Tomé utilizando como base a imagem gerada pelo sensor LISS III a bordo do satélite IRS-P6.

A Tabela 4 ilustra a matriz de confusão gerada na classificação da árvore de decisão. Conforme já citado, três mil pontos de verdades terrestres foram distribuídos aleatoriamente na área de estudo, sendo que todas as amostras de referência que caíram dentro da área da sub-bacia foram utilizadas na validação.

Com exceção das classes corpo d'água e cana-de-açúcar, todas as classes foram contempladas com mais de 50 pontos. Tal fato ocorreu em decorrência da classe corpo d'água ter pouca expressão de área e a cana-de-açúcar ocupar, de modo mais intenso, apenas uma dada área dentro da sub-bacia, o que reduziu a probabilidade de receber um número mais expressivo de pontos aleatórios. Observou-se também um desequilíbrio bastante grande na distribuição dos pontos em virtude da grande diferença de área ocupada pelas classes mapeadas. A classe que recebeu o maior número de pontos foi café produção (354), seguida pela classe solo exposto (220) e pastagem (152). Em virtude da classe corpo d'água ter pouca expressão espacial, considerou-se que o número de pontos de referência (19) foi suficientemente grande para expressar a variabilidade da classificação.

Tabela 4. Matriz de confusão gerada para a classificação por mineração de dados no GeoDMA. A média da acurácia (ACC%) foi de 75%. Já a média da confiabilidade em porcentagem foi de 68% e a média global da diagonal principal (exatidão global) foi de 73%.

Classificação→ Referências↓	Café Formação	Café Produção	Cana	Pastagem	Solo Exposto	Mata	Corpo d'Água	Total	ACC%
Café Formação	57	13	2	0	0	7	0	79	72%
Café Produção	8	219	48	37	7	35	0	354	61%
Cana	0	0	34	2	1	0	0	37	91%
Pastagem	0	5	1	106	38	2	0	152	69%
Solo Exposto	0	1	2	29	188	0	0	220	85%
Mata	2	13	1	2	0	108	0	126	85%
Corpo d'Água	0	3	0	1	0	3	12	19	63%
Total	67	254	88	177	234	155	12	987	
Confiabilidade %	85%	86%	38%	59%	80%	69%	100%		

Analisando a matriz de confusão observa-se que, mesmo com uma grande variabilidade na classificação, os resultados de exatidão (proporção de acertos contidos ao longo da linha) e de confiabilidade (proporção de acertos contidos ao longo da coluna) do mapeamento foram resultados satisfatórios. A classe que teve a maior exatidão individual foi a cana-de-açúcar (91%), seguida pelo solo exposto (85%) e mata (85%). Este valor de exatidão da cana-de-açúcar foi ocasionado pela baixa confusão com outras classes. Para as classes solo exposto e mata houve um pouco de confusão entre solo exposto e pastagem seca e entre mata e café produção. As demais classes tiveram exatidão inferior a 80%.

No caso do café, vários estudos já realizados mostram a dificuldade em separar cultivos de café de outros usos em decorrência de sua alta diversidade espectral (MOREIRA et al., 2004, 2008; DO PRADO et al., 2013). Dependendo do estágio fenológico do cultivo, idade, manejo e condições topográficas, o comportamento espectral do café pode ser semelhante à mata (café produção), ao pasto verde ou cana-de-açúcar (café formação ou café produção em terrenos declivosos) ou pastagem (café velho ou podado).

Observa-se que no geral, os valores de confiabilidade foram menores que os de exatidão. Entretanto, a classe corpo d'água foi a que apresentou a maior confiabilidade (100%), seguido pelo café produção (86%) e café formação (85%). A classe de menor confiabilidade foi a cana-de-açúcar (38%) em decorrência principalmente da confusão com o café produção. A média da exatidão da classificação foi de 75% e a média da confiabilidade foi de 68%. A exatidão global que expressa à proporção de pontos corretamente classificados com relação ao total de pontos amostrados alcançou um valor de 73%. A outra referência utilizada para a avaliação da classificação foi o índice Kappa, que alcançou o valor de 0,66 – sendo classificado na categoria de desempenho “muito bom” como mostrado abaixo na escala de valores da Tabela 5 (LANDIS; KOCH, 1977).

Tabela 5. Valores do índice Kappa e o seu grau de concordância, baseado nesses valores.

Valor de Kappa	Concordância
< 0,00	Péssima
0,00 – 0,20	Ruim
0,20 – 0,40	Razoável
0,40 – 0,60	Boa
0,60 – 0,80	Muito Boa
0,80 – 1,00	Excelente

Fonte: Adaptado de Landis; Koch, (1977).

O índice Kappa expressa à medida da diferença entre a concordância dos dados de referência e a classificação automática, e a probabilidade de concordância entre os dados de referência e a classificação aleatória (FRANÇA; SANO, 2011).

6. DISCUSSÃO

A segmentação com 2499 polígonos foi o melhor resultado obtido. Porém, para se chegar a esse resultado satisfatório, foi necessário escolher um método eficiente. O método utilizando então foi o método de Baatz et al. (2000). Esse método foi eficiente por utilizar a multirresolução e o crescimento por região para segmentar a imagem. No resultado final, a combinação dessas duas técnicas gerou uma imagem segmentada, pronta para ser usada nos passos seguintes de extração e classificação. Sendo a região muito heterogênea em seus tipos de uso da terra, a multirresolução reduz esta heterogeneidade e o crescimento por região, trabalhando não com pixels isolados, mas com regiões, agrupou as mesmas de modo mais eficiente. No entanto, vale resaltar aqui que esse processo evidencia ser basicamente no que diz respeito à divisão da imagem em diversos polígonos durante o processo de segmentação, não afetando diretamente no processo de classificação da imagem.

O algoritmo C4.5 e a técnica de mineração de dados foram eficientes neste trabalho, pois, respectivamente trabalhando com atributos espectrais, espaciais e atributos do terreno (hipsometria, declividade e orientação das vertentes), mais a utilização do aprendizado por máquina, ampliaram o número de variáveis a serem utilizadas no processo de classificação.

A interação intérprete/*software* possibilita uma maior flexibilidade na geração das árvores de decisão e das classificações. O conhecimento da área de estudo por parte do usuário auxilia de modo positivo na hora da coleta de amostras e também na distribuição de pontos de verdades terrestres. As visitas a campo, apesar de darem uma percepção mais limitada da área de estudo, principalmente quando a área é muito grande, ainda assim ajudam a confirmar aquilo que é feito em laboratório, utilizando as imagens.

O resultado obtido por essas técnicas pode ser visto na árvore de decisão da Figura 14 gerada pelo GeoDMA, que reúne os melhores valores para cada classe classificada no mapa temático final da Figura 15. Entretanto, nenhuma propriedade espacial ou topográfica foi selecionada pelo programa na construção da árvore de decisão apresentada.

Em relação a essa classificação exposta na árvore de decisão, o primeiro ponto a se ressaltar aqui, é que apesar de haver na árvore apenas atributos

espectrais, a mineração dos dados dentro do programa é baseada em todos os atributos espectrais e espaciais presentes no processo de extração e classificação. O segundo ponto é que todas as árvores que utilizaram algum atributo espacial tiveram problemas de superestimação de classes na classificação final; claro que isso não significa que o atributo espacial influenciou neste erro.

Durante os vários testes realizados, algumas árvores selecionaram atributos espaciais. Foram selecionados o perímetro, a área e a área-caixa, que analisam as partes internas dos objetos e, a largura e o comprimento, que analisam as medidas dos objetos (Quadro 2). Todavia, nenhuma delas selecionou atributos topográficos. As classes solo exposto e pastagem foram as que mais frequentemente selecionaram os atributos espaciais mencionados; porém, em todas as classificações observou-se superestimava das classes mencionadas. Uma análise visual foi feita para verificar os possíveis acertos e erros de cada classificação.

Quadro 2. Características espaciais baseados em segmentação, adaptado de Korting et al. (2013).

Atributo espacial	Significado
Perímetro	É a quantidade de pixels na borda do objeto.
Área	Retorna a área do objeto. Quando medida em pixels é igual a N .
Área-Caixa	Retorna a área da caixa delimitadora de um objeto, medida em pixels.
Largura	É a largura da caixa delimitadora do objeto.
Comprimento	É a altura da caixa delimitadora do objeto.

Analisando um desses casos, o solo exposto e a pastagem foram classificados no atributo espacial largura. Ao se observar esta classificação foi possível notar que, apesar das duas classes terem sido classificadas neste atributo, houve erro já que houve superestimação, principalmente da pastagem. Baseando-se apenas na largura, a pastagem ocupou grande parte de sua área de modo correto; no entanto, ocupou ainda não apenas áreas de solo exposto, mas também áreas de café produção já podadas e áreas de cana-de-açúcar em início de colheita. O solo exposto foi classificado também no atributo espectral moda e ficou bem distribuído ao longo da área. Entretanto, também ocupou algumas áreas de café produção já podadas e áreas de cana-de-açúcar em início de colheita.

No que diz respeito à distribuição dos tipos de uso do solo e da cobertura vegetal ao longo da área de estudo, vale ainda ressaltar algumas observações relevantes: (a) a ocupação do café na área ao longo de toda a sub-bacia vem

apenas confirmar a importância que essa cultura agrícola tem na região, pois o sul de Minas é o maior produtor de café do país, como já citado; (b) pelas características da região do sul de Minas (mar de morros), o novo Código Florestal (Lei nº 12.651/ 2012) prevê a manutenção de atividades agrossilvipastoris já consolidadas em vertentes e topos de morro nos imóveis rurais antes de 22 de julho de 2008. No entanto, a nova lei proíbe que novas áreas sejam ocupadas por edificações e atividades agrossilvipastoris; e (c) sendo o objetivo deste trabalho testar a classificação por mineração de dados, não fazendo aqui uma análise multitemporal do uso do solo e da cobertura vegetal na área de estudo, houve então excesso aparente de solo exposto. Isso se explica possivelmente pela compatibilidade da imagem com a época de colheita, tanto e principalmente da cana-de-açúcar que pouco aparece na área de estudo, quanto do café. Outra razão deste excesso também está relacionada, mesmo que com um efeito menor, a afloramentos rochosos ao longo da área da sub-bacia.

Após vários testes para se produzir a matriz de confusão da mineração de dados, os valores resultantes na matriz da Tabela 4 exposta acima foram os melhores valores alcançados. Isso se justifica principalmente pela grande heterogeneidade da área de estudo, como observado no mapa de localização (Figura 1) e também no mapa final de uso da terra e de cobertura vegetal (Figura 13), não havendo aqui, como já ressaltado, ação da homogeneidade gerada pela multirresolução do método de segmentação baseado em Baatz et al. (2000). Então, esta alta heterogeneidade e a mistura de classes limitaram os valores de acurácia dentro da matriz.

Desse modo, fica clara a necessidade de mais estudos para se melhorar a acurácia nestes casos específicos de alta heterogeneidade e de mistura de classes. O próprio Korting et al. (2009) aponta a necessidade de otimizar os algoritmos inclusos no sistema do GeoDMA, além da inclusão de ferramentas para a seleção de atributos e manipulação de dados temporais para a análise da evolução dos padrões.

O valor alcançado de 0,66 (muito bom) pelo índice Kappa para a classificação por mineração de dados foi considerado um resultado bom já que na tabela de valores que vai de 0 a 1, seu resultado ficou entre 0,60 – 0,80, ou seja, mais próximo do grau de concordância “excelente”, que vai de 0,80 – 1,00.

7. CONCLUSÃO

A técnica de mineração de dados mostrou-se eficiente no mapeamento do uso da terra e cobertura vegetal em área com grande diversidade de uso como no sul de Minas Gerais. As vantagens apresentadas neste trabalho, no que diz respeito a esta técnica, é que ela, ao permitir o uso de outras variáveis que não só as espectrais, faz com que cada segmento seja minuciosamente analisado e classificado. Além disso, ao se trabalhar com grandes bancos de dados com esta técnica, chega-se à conclusão de que ela é muito eficiente para trabalhos atuais.

No entanto, neste trabalho, apesar de se considerar na mineração de dados os atributos espectrais, espaciais e topográficos, apenas informações espectrais foram selecionadas na classificação, o que sugere pouca relevância de outros atributos no processo de decisão. Todas as bandas do sensor LISS III (bandas do verde, vermelho, infravermelho próximo e infravermelho médio) foram selecionadas no processo de mineração de dados. Tal resultado reforça o potencial das imagens LISS-III na identificação e classificação do uso da terra e cobertura vegetal.

A dificuldade maior de se usar a mineração de dados, mas que também se converte em vantagem são as frequentes mudanças que o intérprete faz em relação aos atributos espaciais e espectrais presentes no programa, já que os mesmos podem interferir de modo negativo na classificação final. A dificuldade então se resume ao grande número de testes que precisa ser feito para se alcançar uma boa classificação; já a vantagem é que estes testes, que são analisados visualmente pelo intérprete, possibilitam um maior controle dos atributos que serão usados pelo programa para gerar uma classificação ideal para o usuário.

O *software* GeoDMA usado aqui foi eficiente. Esta eficiência ao utilizá-lo ocorreu principalmente pela sua facilidade de manipular dados e informações, apesar de trabalhar com uma técnica complexa como a de mineração de dados. O seu uso como um *plugin* do *TerraView* permite que se tenha uma interface de fácil interpretação, já que as próprias ferramentas que são usadas neste *software* são de fácil manipulação, ajudando e facilitando o trabalho do usuário, principalmente no que diz respeito ao tempo gasto no processamento da imagem.

A maior dificuldade em relação ao GeoDMA se deu pelo domínio das ferramentas e etapas de análise de imagem presentes neste programa, já que o primeiro contato com ele foi feito a partir deste trabalho. Contudo, ao se adquirir um

pouco mais de contato com o programa, a sua utilização se deu de forma fácil e rápida.

Áreas muito heterogêneas ainda pode ser um problema, pois, o grande parcelamento do solo e a mistura de classes parecem que ainda não permitem índices de acurácia maiores. Sendo assim, apesar das vantagens mostradas aqui, ainda há de se fazer mais estudos e testes para se aproveitar todo o potencial que essa técnica permite alcançar.

Os resultados alcançados neste trabalho só reforçam ainda mais que o sensoriamento remoto aplicado ao uso da terra e cobertura vegetal, ao utilizar as técnicas de classificação de imagens como a mineração de dados, pode e deve ser auxílio e ponto de partida para um bom planejamento ambiental e agrícola.

AGRADECIMENTOS

Os autores expressam os seus agradecimentos ao Programa de Pós-Graduação em Ecologia e Tecnologia Ambiental e ao Laboratório de Geoprocessamento do curso de Geografia da Unifal-MG. Agradecem também a CAPES pelo apoio financeiro concedido, ao Dr. Thales Sehn Korting e ao Rodrigo César Pereira Nicolau.

REFERÊNCIAS BIBLIOGRÁFICAS

AB'SÁBER, A. N. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. 2ª ed. São Paulo: Ateliê Editorial, 2003. 160 p.

BAATZ, M. et al. **User Guide 4 (eCognition)**. Definiens Imaging. Copyright © 2000 - 2004. Germany, 2004. 485 p.

BAATZ, M.; SCHÄPE, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. **Journal of Photogrammetry and Remote Sensing**. Heidelberg, v. 58, n. 3-4, p. 12-23, 2000.

BERNARDES, T. et al. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. **Remote Sensing**, n. 9, p. 2492-2509, 2012, doi: 10.3390/rs4092492.

BINS, L. S. et al. Satellite Imagery Segmentation: a region growing approach. In: VIII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 1996, Salvador, Brasil.

Anais VIII Simpósio Brasileiro de Sensoriamento Remoto, Salvador: INPE, p. 677- 680, 14-19, abril, 1996.

BLASCHKE, T. Object based image analysis for remote sensing. **International Society for Photogrammetry and Remote Sensing (ISPRS)**, 65, p. 2-16, 2010.

BURROUGH, P. A. **Principles of Geographical Information Systems for Land Resources Assessment**. Oxford: Clarendon, 1986.

BURROUGH, P. A.; MCDONNELL, R. A. **Principles of Geographical Information Systems**. OXFORD UNIVERSITY PRESS, 1998.

CABRAL, P. **Erros e Incertezas nas Medições**. IEP/ ISEP, 2004, p. 116.

CÂMARA, G. et al. **Análise Espacial de Áreas**. In: Druck, S.; Carvalho, M. S.; Câmara, G.; Monteiro, A. V. M. (eds.) "Análise Espacial de Dados Geográficos". Brasília, EMBRAPA, 2004.

CARREIRAS, J. M. B.; SHIMABUKURO, Y. E.; PEREIRA, J. M. C. Fraction images derived from SPOT-4 Vegetation data to assess land-cover change over the State of Mato Grosso, Brazil. **International Journal of Remote Sensing**. V. 23, n. 23, p. 4979-4983, 2002.

CARTAS TOPOGRÁFICAS. **Manual Técnico de Uso da Terra**, nº 7. In: Manuais Técnicos em Geociências. 2ª Ed. IBGE, 2006. 91 p.

CERON, A. O.; DINIZ, J. A. F. O uso de fotografias aéreas na identificação das formas de utilização agrícola da terra. **Revista Brasileira de Geografia**. N. 2, p.161-173, 1966.

CHUBEY, M. S.; FRANKLIN, S. E.; WULDER, M. A. Object-based Analysis of Ikonos-2 Imagery for Extraction of Forest Inventory Parameters. **Photogrammetric Engineering & Remote Sensing**. V. 72, n. 4, p. 383–394, 2006.

CONGALTON, R. G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. **Remote Sens. Environ.** 37, p. 35-46, 1991.

CONGALTON, R. G.; GREEN, K. A Practical Look at the Sources of Confusion in Error Matrix Generation. **Photogrammetric Engineering & Remote Sensing**. V. 59, n. 5, p. 641-644, 1993.

CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Ed. rev. 4ª Reimpressão. Campinas, SP: IG/UNICAMP, 2002. 170 p.

D'ALGE, J. C. L. **Geoprocessamento - Teoria e Aplicações - Parte I - Cap. 6 - Cartografia para Geoprocessamento**. INPE - Instituto Nacional de Pesquisas Espaciais, 2001. Disponível em: <www.dpi.inpe.br/gilberto/livro/introd/>. Acesso em: 10 maio 2013.

DESMET, P. J. J. Effects of interpolation errors on the analysis of DEMs. **Earth Surface Processes and Landforms**. V. 22, p. 563-580, 1997.

DGI (Divisão de Geração de Imagens) - INPE. **Características do satélite IRS-P6 ou RESOURCESAT-1/ sensor LISS III.** Disponível em: <http://www.dgi.inpe.br/CDSR/manage.php?INDICE=P6LIS333109320100731&DONT_SHOW=0>. Acesso em: 20 fev. 2013.

DO PRADO, D. F. C. et al. Mapeamento do cultivo de café no sul de Minas Gerais utilizando imagens Landsat-5 TM. In: XVI SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO - SBSR, 2013, Foz do Iguaçu, PR, Brasil. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Foz do Iguaçu, PR, Brasil, INPE: 13-18, abril, 2013.

DPI (Divisão de Processamento de Imagens) - INPE. **GeoDMA - Geographic Data Mining Analyst.** Disponível em: <<http://www.dpi.inpe.br/menu/Projetos/geodma.php>>. Acesso em: 15 mar. 2013.

EMBRAPA – Monitoramento por Satélite. **IRS - Indian Remote Sensing Satellite.** Disponível em: <http://www.sat.cnpm.embrapa.br/conteudo/missao_irs.php>. Acesso em: 05 fev. 2013.

EMBRAPA – Brasil maior produtor mundial de café. Disponível em: <<https://www.embrapa.br/web/portal/busca-de-noticias/-/noticia/1477578/brasil-tem-dia-nacional-do-cafe->>. Acesso em: 15 dez. 2013.

EIPHANIO, J. C. N; LEONARDI, L.; FORMAGGIO, A. R. Relações entre parâmetros culturais e resposta espectral de cafezais. **Pesquisa Agropecuária Brasileira**. V. 29, n. 3, p. 439-447, 1994.

ESPINDOLA, G. Emprego dos Índices de Autocorrelação Espacial na Avaliação da Segmentação de Imagens de Sensoriamento Remoto. In: IV COLOQUIO BRASILEIRO DE CIÊNCIAS GEODÉSICAS - IV CBCG, 2005, Curitiba. **Anais IV Colóquio Brasileiro de Ciências Geodésicas - IV CBCG**. Curitiba, 16-20, maio, 2005.

FELGUEIRAS, C. A. **CAP. 7: MODELAGEM NUMÉRICA DE TERRENO.** In: INTRODUÇÃO À CIÊNCIA DA GEOINFORMAÇÃO. Gilberto Câmara, Clodoveu Davis e Antônio Miguel Vieira Monteiro (Orgs.). INPE. São José dos Campos – SP, 2001. 345 p.

FLORENZANO, T. G. **Iniciação em Sensoriamento Remoto.** 3 ed. ampl. e atual. São Paulo: Oficina de Textos, 2011. 128 p.

FLORENZANO, T. G. **Os Satélites e Suas Aplicações.** Série Espacializando. São José dos Campos – SP. SindCT, 2008. 52 p.

FRANÇA, A. M. da S.; SANO, E. E. **Mapeamento de Áreas de Campo Limpo Úmido no Distrito Federal a Partir de Fusão de Imagens Multiespectrais.** Soc. & Nat. Uberlândia, ano 23, n. 2, p. 197-209, maio/ago. 2011.

FURTADO, C. **Formação Econômica do Brasil.** 32ª Ed. São Paulo. Companhia Editorial Nacional, 2005.

GAMA, J. **Árvores de Decisão**. Palestra ministrada no Núcleo da Ciência de Computação da Universidade do Porto, Porto, 2002. Disponível em: <http://www.dcc.fc.up.pt/~lnes/aulas/MIM/arvores_de_decisao.pdf>. Acesso em: 14 ago. 2013.

GONZALEZ, R. C.; WOODS, R. E. (1992). **Processamento digital de imagens**. Tradução Cristina Yamagami e Leonardo Piamonte. 3. ed. São Paulo: Pearson Prentice Hall, 2010. 624 p.

GORTE, B. G. H.; KOOLHOVEN W. 1990. Interpolation between isolines based on the Borgefors distance transform. *ITC Journal*, 3, p. 245-247, 1990.

HALARICK, R. M.; SHAPIRO, L. G. Survey: image segmentation techniques. **Computer, Vision, Graphics, and Image Processing**. V. 29, p. 100-132, 1985.

IBGE. **Censo Agropecuário 2006. Brasil, Grandes Regiões e Unidades da Federação**. IBGE, Rio de Janeiro, 2006. 777 p.

IBGE. **Histórico de ocupação dos municípios que compõem a bacia do rio São Tomé**. Disponível em: <<http://cidades.ibge.gov.br/xtras/uf.php?lang=&coduf=31&search=minas-gerais>>. Acesso em: 20 ago. 2012.

IBGE. **Introdução ao Processamento Digital de Imagens**. Manuais Técnicos em Geociências – n. 9. IBGE. Rio de Janeiro, 2000.

IBGE. **Produção agropecuária dos municípios que compõem a bacia do rio São Tomé**. Disponível em: <<http://www.cidades.ibge.gov.br/xtras/uf.php?lang=&coduf=31&search=minas-gerais>>. Acesso em: 08 out. 2013.

IBGE. **Manual Técnico de Uso da Terra**, n. 7. In: Manuais Técnicos em Geociências. 2ª Ed. IBGE, 2006. 91 p.

INPE. **Estudos Ambientais**. Disponível em: <<http://satelite.cptec.inpe.br/home/novoSite/index.jsp>>. Acesso em: 20 ago. 2013.

JENSEN, J. R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. Tradução: José Carlos Neves Epiphany et al. São Jose dos Campos: Parêntese, 2009. 598 p.

KAWAKUBO, F. S. **Metodologia de Classificação de Imagens Multiespectrais aplicada ao Mapeamento do Uso da Terra e Cobertura Vegetal na Amazônia: Exemplo de Caso na Região de São Felix do Xingu, Sul do Pará**. 2010. 129f. Tese (Doutorado em Geografia Física). FFLCH/ USP. São Paulo, 2010.

KAWAKUBO, F. S.; MORATO, R. G.; LUCHIARI, A. **Técnicas de Sensoriamento Remoto**. In: Luis Antonio Bittar Venturi. (Org.). Geografia: Práticas de Campo, Laboratório e Sala de Aula. São Paulo: Sarandi, 2011, 528 p.

KAWAKUBO, F. S.; MORATO, R. G.; LUCHIARI, A. Use of fraction imagery, segmentation and masking techniques to classify land-use and land-cover types in

the Brazilian Amazon. **International Journal of Remote Sensing**. V. 34, n. 15, p. 5452–5467, 2013.

KORTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA – Geographic Data Mining Analyst. **Computers & Geosciences**, 57, p. 133-145, 2013.

KORTING, T. S. et al. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. In: XIV SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2009, Natal, Brasil. **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto**, Natal, Brasil: INPE, p. 7813-7820, 25-30, abril, 2009.

KUBIK, K.; BOTMAN, A. G. Interpolation accuracy for topographic and geological surfaces. **ITC Journal**, n. 2, p. 136-274, 1976.

LANDIS, J. R.; KOCH G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, 33, p. 159–174, 1977.

Lei nº 12.651/ 2012. **Novo Código Florestal**. Disponível em: <<http://www2.camara.leg.br/legin/fed/lei/2012/lei-12651-25-maio-2012-613076-normaatuizada-pl.pdf>>. Acesso em: 18 maio 2014.

LI, Z. Effects of checkpoints on the reliability of DTM accuracy estimates obtained from experimental tests. **Photogrammetric Engineering and Remote Sensing**. N. 57, p.1333-1340, 1991.

LOPES, A.; MOTA, C. G. **História do Brasil: uma interpretação**. São Paulo: Editora Senac São Paulo, 2008. 1056 p.

LOPES, A. S.; GUILHERME, L. R. G. **Cap. I – Fertilidade do solo e produtividade agrícola**. In: Fertilidade do solo. Roberto Ferreira de Novais, Victor Hugo Alvarez V., Nairam Félix De Barros, Renildes Lúcio Fontes, Reinaldo Bertola Cantarutti e Júlio César Lima Neves (Editores). Sociedade Brasileira de Ciência do Solo (SBCS), Viçosa, 2007. 1017 p.

MACHADO, C. A. S.; QUINTANILHA, J. A. **Módulo de treinamento: Sistemas de Informações Geográficas (SIG) e Geoposicionamento: uma aplicação urbana**. Escola Politécnica da universidade de São Paulo. Departamento de Engenharia de Transportes. São Paulo, 2008. 113 p.

MANGABEIRA, J. A. de C. et al. **Avaliação do levantamento do uso das terras por imagens de satélites de alta e média resolução espacial**. In: Susceptibilidade à erosão na Região Oeste do Estado da Bahia – Comunicado Técnico, 11. Campinas, SP. Dezembro, 2003.

MAZOYER, M.; ROUDART, L. **História das agriculturas pelo mundo: do neolítico à crise contemporânea**. Tradução: Cláudia F. Falluh Balduino Ferreira. São Paulo: Editora UNESP, Brasília, DF: NEAD, 2010. 568 p.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat-5. **Pesquisa Agropecuária Brasileira**, Brasília, v. 39, n. 3, p.223-231, 2004.

MOREIRA, M. A.; BARROS, M. A.; RUDORFF, B. F. T. Geotecnologias no mapeamento da Cultura do Café em Escala Municipal. **Sociedade & Natureza**, Uberlândia, 20 (1), p. 101-110, junho, 2008.

NIJMEIJER et al. **ILWIS 3.0 Academic user's guide**. Unit Geo Software Development. International Institute for Geo-Information Science and Earth Observation (ITC). Enschede, 2001. 520 p.

OLIVEIRA, J. de A. et al. Extração de atributos de Forma para Classificação de Imagens de alta Resolução do Satélite HRC/CBERS-2B. In: XIV SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2009, Natal, Brasil. **Anais XIV Simpósio Brasileiro de Sensoriamento Remoto**, Natal, Brasil, INPE: p. 7015-7022, 25-30, abril, 2009.

PAL, N. R.; PAL, S. K. A review on image segmentation techniques. **Pattern Recognition**. V. 26, n. 9, p. 1277-1294, 1993.

PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, 11, p. 1633–1644, 2007.

PESSOA, M. C. P. Y.; FERNANDES, E. N. **Modelagem matemática e simulação de sistemas aplicadas ao planejamento ambiental e da atividade agrícola**. In: Planejamento ambiental do espaço rural com ênfase para microbacias hidrográficas: manejo de recursos hídricos, ferramentas computacionais e educação ambiental. Marco Antonio Ferreira Gomes, Maria Conceição Peres Young Pessoa, editores técnicos. – Brasília, DF: Embrapa Informação Tecnológica, 2010. 407 p.

QUINLAN, J. R. **C4. 5: Programs for Machine Learning**. [S.l.]: Morgan kaufmann, 1993.

RADAMBRASIL, **Levantamento dos recursos naturais: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra**. Folhas SF. 23 / 24, Rio de Janeiro/Vitória. Rio de Janeiro: Ministério das Minas e Energia, 1983. 780 p.

RAVIBABU, M. V.; JAIN, K. Digital elevation models accuracy aspects. **Journal of Applied Sciences**. V. 8, n. 1, p.134-139, 2008.

ROSS, J. S. L. (org.) **Geografia do Brasil**. 5 ed. rev. e ampl. – São Paulo: Editora da Universidade de São Paulo, 2005.

RUFINO, J. L. dos S. et al. **Introdução e Metodologia de Estudo**. In: Caracterização da Cafeicultura de Montanha de Minas Gerais / coordenadores, Pierre Santos Vilela, José Luis dos Santos Rufino. – Belo Horizonte: INAES, 2010. 300 p.

SEGANTINE, P. C. L. **Estudo do Sinergismo entre os Sistemas de Informação Geográfica e o de Posicionamento Global**, 2001. 237f. Livre-Docência. Escola de Engenharia de São Carlos, Departamento de Transporte/ USP. São Carlos-SP, 2001.

SILVA, A. C. et al. Desenvolvimento sustentável da bacia hidrográfica do rio São Tomé. *Gestão de Extensão e Estudos Independentes: Revista do I Fórum de Extensão Universitária*, 2011. 7 p.

SILVA, B. C. da *et al.* Comparação entre dados dos sensores LISS-III/ Resourcesat-1 e ETM+/ Landsat 7. In: XVI SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO - SBSR, 2013, Foz do Iguaçu, PR, Brasil. **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR**, Foz do Iguaçu, PR, Brasil, INPE: p. 9138-9142, 13-18, abril, 2013.

SILVA, M. P. S. et al. Remote-sensing image mining: Detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, 29(16), p. 4803–4822, 2008.

SINDICAFÉ – MG (Sindicato da Indústria de Café do Estado de Minas Gerais). **Produção de café no Sul de Minas Gerais**. Disponível em: <<http://sindicafe-mg.com.br/plus/modulos/conteudo/?tac=tipos-de-cafe>>. Acesso em: 10 maio 2014.

SOARES, V. P. et al. Mapeamento das Áreas de Preservação Permanente e dos Fragmentos Florestais Naturais como Subsídio à Averbação de Reserva legal em Imóveis rurais. **Cerne**, Lavras, v. 17, n. 4, p. 555-561, out./dez. 2011.

SPÖRL, C.; ROSS, J. L. S. Análise Comparativa da Fragilidade Ambiental Com Aplicação de Três Modelos. **GEOUSP – Espaço e Tempo**, São Paulo, n. 15, p. 39-49, 2004.

STRAHLER, A. N. Hypsometric (area-altitude) – analysis of erosion al topography. **Geological Society of America Bulletin**, v. 63, n. 10, p. 1117-1142, 1952.

TIAN, J.; CHEN, D. M. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. **International Journal of Remote Sensing**. V. 28, n. 20, p. 4625–4644, 2007.

VEIGA, L. A. K.; ZANETTI, M. A. Z.; FAGGION, P. L. **Fundamentos de Topografia**. UFPR, Paraná, 2012. 288 p.

VIEIRA, M. A. et al. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. **Remote Sensing of Environment**, 123, p. 553-562, 2012.

YAMAMOTO, J. K. A review of Numerical Methods for the Interpolation of Geological Data. **Anais da Academia Brasileira de Ciências**. V. 70, n.1, p. 92-116, 1998.

APÊNDICE A – Artigo submetido

CLASSIFICAÇÃO DO USO DA TERRA E COBERTURA VEGETAL UTILIZANDO MINERAÇÃO DE DADOS: EXEMPLO NO SUL DE MINAS GERAIS

Land-use and land-cover classification using data mining: example in southern Minas Gerais

RODRIGO CESÁRIO JUSTINO¹
FERNANDO SHINJI KAWAKUBO¹

¹Universidade Federal de Alfenas/ MG – UNIFAL/ MG
Alfenas – MG – Brasil

rodrigofriend@bol.com.br; fsk@gmail.com

RESUMO

O objetivo deste estudo é realizar a classificação do uso da terra e cobertura vegetal na bacia do Rio São Tomé no sul do estado de Minas Gerais utilizando imagens multiespectrais geradas pelo sensor LISS III (*Linear Imaging Self-Scanner*) a bordo do satélite IRS (*Indian Resource Satellite*) e técnicas de mineração de dados. A área de estudo localiza-se em uma região de grande tradição de cultivo de café cujas propriedades são predominantemente de pequeno e médio portes. Para realizar a classificação das imagens, foi utilizado o aplicativo GeoDMA (*Geographic Data Mining Analyst*) que possui algoritmos para segmentação, extração de atributos, seleção de feições e classificação. Os resultados mostraram o grande potencial das técnicas de mineração de dados na classificação de imagens de satélite. Algumas das vantagens da utilização de mineração de dados incluem: a possibilidade de incorporar no processo de classificação um grande número de variáveis, sejam espectrais, espaciais e atributos do terreno; a geração de classificações consistentes; e a simplicidade para interpretar a suas estruturas de classificação.

Palavras-chave: mineração de dados, sensoriamento remoto, GeoDMA, segmentação, árvore de decisão, exatidão.

ABSTRACT

The objective of this study is to classify the land use and land cover of the São Tomé river watershed located in the south of the Minas Gerais state by using multispectral imagery collected by the Linear Imaging Self-Scanner onboard the Indian Resource Satellite and data mining techniques. The study area lies in a region with great tradition of coffee cultivation whose properties are predominantly of small and medium sizes. To carry out the image classification, we used the Geographic Data Mining Analyst (GeoDMA), a toolbox that has algorithms for segmentation, feature extraction, feature selection and classification. The results showed the great potential of the data mining techniques for remote sensing imagery classification. Some advantages encompass the possibility of incorporate in the classification process a great variety of informations that includes spectral, spatial and topographic attributes, the generation of high classification accuracy, and the simplicity to interpret their classification structure.

Keywords: data mining, remote sensing, GeoDMA, segmentation, decision tree, accuracy.

1. INTRODUÇÃO:

A tecnologia do sensoriamento remoto (SR) constitui em uma das principais fontes de informações de apoio aos estudos destinados ao planejamento agrícola e ambiental. Além de fornecer uma visão panorâmica da paisagem, o SR permite a coleta de dados em intervalos de tempos regulares (devido à periodicidade de visita do satélite), em regiões do espectro eletromagnético inacessíveis à visão humana (em razão do caráter multiespectral dos sensores), com confiabilidade e com baixo custo relativo.

Uma das principais aplicações do SR voltadas para o planejamento agrícola e ambiental é o mapeamento do uso da terra e cobertura vegetal. Tradicionalmente, este tipo de mapeamento é feito utilizando os elementos de reconhecimento da fotogrametria interpretativa (CERON; DINIZ, 1966). Todavia, tal metodologia adota critérios bastante subjetivos no processo de decisão, além de exigir bastante tempo na sua execução, o que encarece o custo do mapeamento.

As técnicas de classificação digital de imagens de satélite têm se sofisticado ao longo dos anos com significativas melhorias no processo de reconhecimento de padrões de uso da terra e cobertura vegetal. Uma das abordagens mais promissoras é a classificação de imagens orientada a objeto. A classificação orientada a objeto tem como unidade de análise grupos de pixels delimitados por segmentos - ao invés de pixels isolados como feita na classificação tradicional. Esses segmentos geralmente incorporam um amplo conjunto de informações que inclui atributos espectrais, texturais e espaciais (BLASCHKE, 2010) que são usados para treinar ou construir uma base de conhecimento para a classificação de imagens (KAWAKUBO et al., 2013).

A classificação baseada no conhecimento simula o processo de fotointerpretação. Assim, critérios espectrais, texturais e de contexto são considerados no processo de decisão. De acordo com Vieira et al. (2012), uma etapa importante desta abordagem é a construção do modelo de interpretação de imagem (conhecimento). Neste modelo, o analista deve ter uma exata noção dos

melhores atributos a serem utilizados na classificação. Condição esta, muitas vezes não atendida em virtude da dificuldade em gerenciar a grande quantidade de informações disponíveis e do nível de detalhamento exigido.

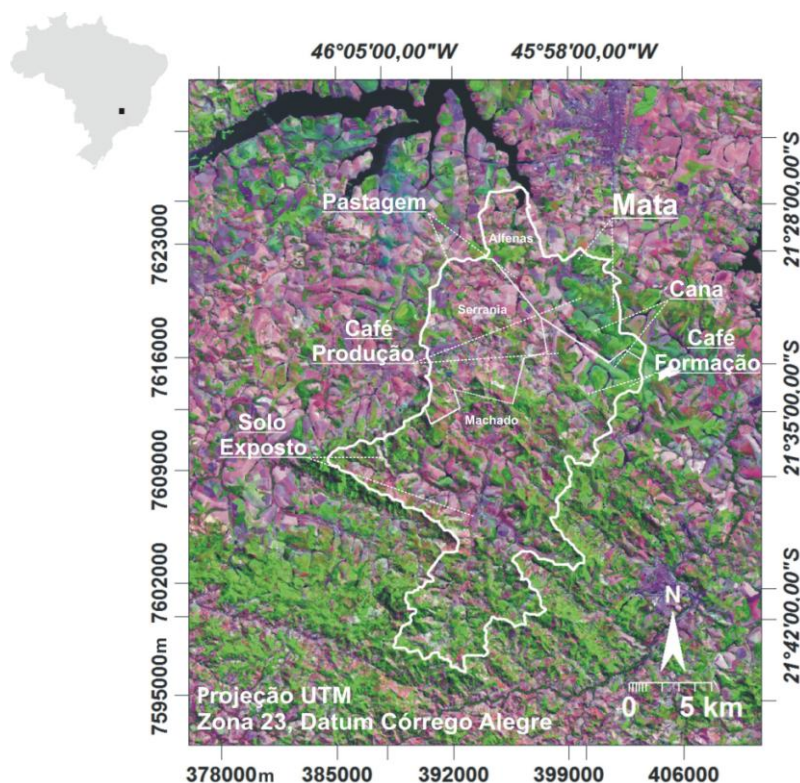
Como forma de superar tais limitações, a adoção de técnicas de mineração de dados surge como uma alternativa viável para geração automática de uma estrutura de conhecimento (SILVA et al., 2008; VIEIRA et al., 2012). A adoção de técnicas de mineração de dados em classificação de imagens de satélite ainda é pouco disseminada e necessita de estudos mais aprofundados como forma de avaliar o potencial e a limitação da técnica. Em vista dessas demandas existentes, o presente trabalho tem como objetivo utilizar a técnica de mineração de dados aplicada ao mapeamento do uso da terra e cobertura vegetal no sul de Minas Gerais. Atributos espectrais, texturas, espaciais e topográficos são explorados no processo de classificação.

2. ÁREA DE ESTUDO:

A área de estudo corresponde à bacia do rio São Tomé localizado no sul de Minas Gerais (Figura 1). O rio São Tomé constitui um importante afluente do reservatório de Furnas, drenando uma área de aproximadamente 34 mil hectares compreende parte dos municípios de Machado, Serrania e Alfenas (SILVA et al., 2011). A nascente do rio São Tomé localiza-se no município de Machado a cerca de 1200 metros de altitude. A geomorfologia da região está incluída na unidade morfoclimática do domínio das regiões serranas de morros mamelonares caracterizadas por morros arredondados em formato de “meia-laranja” (AB’ SABER, 1963).

O clima é subtropical classificado como Cwa segundo Köppen (PEEL et al., 2007), com verão úmido e inverno seco. A vegetação típica da região é formada por floresta estacional semidecidual ombrófila mista e cerrado (RADAMBRASIL, 1983). Com relação ao uso da terra, a região destaca-se pelo plantio do café e atividade pecuária. Mais recentemente, observa-se também o avanço crescente na região do plantio da cana-de-açúcar.

Figura 1 - Mapa de localização da bacia do rio São Tomé, sul de Minas Gerais. A imagem corresponde a uma composição colorida do sensor IRS-P6 utilizando as bandas do infravermelho médio (b5), infravermelho próximo (b4) e vermelho (b3) combinadas em RGB. A imagem foi adquirida em 31 de julho de 2010 correspondente a órbita/ponto 331/93.



3. MATERIAIS E MÉTODO

3.1 Derivações de Informações do Terreno

Foram utilizadas como base cartográfica quatro cartas topográficas na escala 1: 50000 publicadas pelo IBGE, folhas: Areado (FOLHA SF-23-V-D-I-4), Alfenas (FOLHA SF-23-I-I-3), Campestre (FOLHA SF-23-V-D-IV-2) e Machado (FOLHA SF-23-I-III-1). Utilizando as curvas de nível e pontos cotados, foi gerado um Modelo Digital do Terreno (MDT), que consiste numa representação matemática contínua que descreve as variações relacionadas à altitude do terreno. Neste trabalho, optou-se pela geração do MDT utilizando um interpolador linear que se baseia na distância de Borgefors. A geração deste modelo é dividida em duas etapas. A primeira é a conversão das linhas de contorno do formato vetorial para o raster. Após a conversão, os pixels cobertos pela linha de contorno assumem os valores de altitude das linhas correspondentes e os demais pixels possuem

valores indefinidos. Na segunda etapa é realizada a interpolação linear feita entre os pixels com valores de altitude para obter as elevações dos pixels com valores indefinidos. A partir do MDT foram derivados os seguintes produtos: informações de hipsometria, declividade, orientação das vertentes e hierarquia de sub-bacias hidrográficas seguindo a metodologia de Strahler (1952).

3.2 Imagens de Satélite e Pré-processamento

A imagem utilizada neste trabalho foi obtida junto ao INPE e foi gerada pelo sensor LISS III (*Linear Imaging Self-Scanner*) que está a bordo do satélite indiano IRS - P6 (*Indian Resource Satellite*) ou RESOURCESAT-1. O LISS III é um sensor do tipo *Pushbroom* que possui quatro faixas espectrais que operam nos intervalos do verde (0.52-0.59 μm (b2)), vermelho (0.62-0.68 μm (b3)), infravermelho próximo (0.77-0.86 μm (b4)) e infravermelho médio (1.55-1.70 μm (b5)). A imagem possui resolução espacial de 23,5 metros no terreno quantificado em 7 bits e com tempo de revisita de 24 dias (EMBRAPA, 2013).

As especificações do sensor LISS III são compatíveis com outros sistemas de resolução moderada como as do Landsat -5 (*Thematic Mapper - TM*), -7 (*Enhanced Thematic Mapper Plus - ETM+*), -8 (*Operational Land Imager - OLI*) e CBERS (*High Resolution CCD Cameras*), o que permite comparação das informações extraídas ao longo do tempo. Tal característica justifica a importância de sua utilização em um contexto de complementaridade de informações.

A primeira etapa adotada no tratamento das imagens consistiu na correção geométrica da imagem, que tem como objetivo corrigir as distorções geométricas causadas principalmente pela rotação da Terra e da instabilidade da plataforma (CRÓSTA, 2002). O procedimento de correção geométrica também é um passo fundamental para permitir a integração das imagens com outras fontes cartográficas e possibilitar a comparação de imagens ao longo dos anos e entre instrumentos diferentes. A correção geométrica foi realizada por meio da identificação de pontos de controle na imagem utilizando como referência a base cartográfica definida no projeto. Foram selecionados 15 pontos de controle (*ground control point - GCP*) distribuídos ao longo de toda a área de estudo. O erro de georreferenciamento utilizado a transformação afim foi inferior a 1,0 pixel.

Em seguida, as imagens foram reamostradas com o interpolador espacial vizinho mais próximo. A vantagem da utilização do vizinho mais próximo em relação a outros interpoladores, como o bilinear e o bicúbico é que o algoritmo preserva os valores dos pixels da imagem original. Condição esta importante para a etapa de classificação multiespectral. O espaçamento da grade definido no momento da reamostragem foi de 20 metros, resultando no final deste processo, uma imagem com dimensão de 1717 linhas e 1454 colunas.

3.3 Análise das Imagens e Mineração

Esta etapa foi realizada utilizando o aplicativo GeoDMA (*Geographic Data Mining Analyst*), desenvolvido por Korting et al. (2009; 2013). O GeoDMA integra métodos de análise de imagens com técnicas de mineração de dados (DPI/INPE, 2011). Segundo Korting et al. (2013) o GeoDMA funciona como um *plugin* do TerraView e possui todas as fases de análise de imagens como segmentação, extração e seleção de atributos, mineração de dados, análise multitemporal etc.

Segundo Bins et al. (1996), o processo de segmentação é uma das primeiras etapas na análise de imagens de SR onde a imagem é dividida em regiões que melhor representam os objetos presentes na cena. Na abordagem por segmentação, ao invés de trabalhar a imagem no nível do pixel, trabalha-se com o conceito de objetos, ou seja, conjuntos de pixels agregados que possui alguma propriedade em comum, como cor, textura, padrão etc.

Para realizar a segmentação, adotou-se o algoritmo por crescimento de região baseado em Baatz et al. (2000) que incorpora o conceito de multiresolução. A multiresolução consiste em minimizar a heterogeneidade dos objetos de uma imagem (BAATZ et al., 2004; TIAN; CHEN, 2007), incorporando objetos menores dentro de objetos maiores com o objetivo de reduzir a heterogeneidade espacial (CHUBEY et al., 2006). No GeoDMA três critérios são definidos: 1) compactidade, que trabalha com a densidade dos objetos compactados e não compactados dentro da imagem; 2) escala, que influência no tamanho médio dos objetos; e 3) cor, que tem relação com as propriedades espectrais das bandas (BAATZ et al., 2004).

Procurou-se um resultado de segmentação que não houvesse nem falta e nem excesso de segmentos. A análise dos resultados da segmentação foi feita de maneira visual verificando se os segmentos foram bem delimitados e ajustados ao conteúdo da informação presente na imagem (KAWAKUBO et al., 2009; KAWAKUBO et al., 2013).

Realizada a segmentação, iniciou-se a etapa de extração dos atributos. Além das bandas multiespectrais do sensor LISS III (bandas do verde (b2), vermelho (b3), infravermelho próximo (b4) e infravermelho médio (b5)), foram também incluídas as informações derivadas do MDT (hipsometria, declividade, orientação das vertentes e a hierarquia de bacia hidrográfica). As informações topográficas foram incluídas porque observações em campo constataram um forte controle do relevo no ordenamento territorial da bacia.

Estudos realizados por do Prado et al. (2013) na região de Muzambinho e Cabo Verde (nas proximidades a Alfenas, MG) também mostraram uma forte influência do relevo na distribuição espacial do plantio do café. Os autores observaram que o plantio do café distribui-se entre 800 - 1250 metros de altitude com predomínio na cota de 1000 metros. A declividade também exerce grande influência, sendo o café plantado normalmente da meia vertente em direção aos topos, evitando-se os fundos de vale em função dos maiores riscos de geada.

Além das propriedades espectrais e topográficas, também foram extraídas propriedades texturais e espaciais de cada objeto delimitado pela segmentação. As propriedades texturais são importantes porque incorpora os conceitos de variabilidade interna das classes. Já as propriedades espaciais são importantes porque as formas dos segmentos fornecem fortes indícios dos tipos de uso e de cobertura vegetal. As formas regulares, por exemplo, correspondem a diferentes plantios e as formas irregulares, as coberturas vegetais naturais.

Após a extração de atributos, iniciou-se a fase de amostragem. Foram selecionadas amostras de 7 classes de uso da terra e cobertura vegetal: Corpo d'água, Café formação, Café produção, Cana-de-açúcar, Mata, Pastagem e Solo exposto (esta classe também inclui os afloramentos rochosos). O Café formação corresponde aos plantios de café em sua fase mais jovem, enquanto o Café produção inclui os plantios em estágio de produção de grãos ou os plantios que foram podados.

Para a classificação, utilizou-se o algoritmo C4.5 desenvolvido por Quinlan (1993) para a geração da árvore de decisão. A escolha dos atributos é baseada em uma métrica de entropia denominada de ganho de informação. Os atributos com maiores ganhos de informação são selecionados na construção da árvore de decisão. A utilização de árvores de decisão em classificação de imagens de satélite possui várias vantagens quando comparada com outros classificadores tradicionais, como por exemplo, o da máxima verossimilhança (Maxver). Além de não requerer suposições com relação à distribuição das amostras, a árvore de decisão é de fácil interpretação, possibilita considerar variáveis categóricas e numéricas e permite incorporar relações não lineares entre classes (FRIEDL; BRODLEY, 1997).

No presente estudo, a classificação por árvore de decisão foi realizada de maneira interativa selecionando os atributos de entrada e analisando os acertos e erros contidos na classificação. Foram realizados vários testes de classificação utilizando diferentes atributos. Os atributos indicados na árvore de decisão que causavam erros nas classificações (analisados de maneira qualitativa) foram sendo excluídos em cada teste até que o resultado da classificação atingisse um nível considerado satisfatório.

3.4 Análise de Exatidão

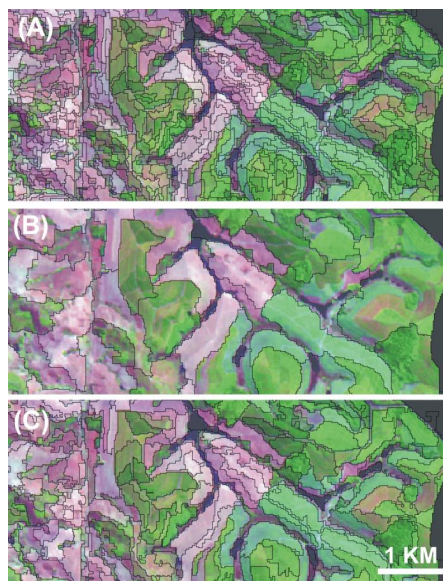
Realizou-se a análise quantitativa da exatidão da classificação por meio de uma matriz de confusão. A matriz de confusão é formada por linhas e colunas onde são confrontados os resultados da classificação com as amostras de referência. Os valores distribuídos na diagonal principal desta matriz correspondem aos acertos obtidos no mapeamento. Já os valores dispostos fora da diagonal principal correspondem às confusões de classificação que podem estar associadas à inclusão ou exclusão de classes (CONGALTON, 1991). Para a construção desta matriz foram coletados 3000 pontos de verdades terrestres distribuídos aleatoriamente na área de estudo, sendo que todas as amostras de referência que caíram dentro da área de estudo foram utilizadas na validação. A checagem destes pontos foi feita com informações obtidas em trabalhos de campo e com auxílio do *Google Earth*.

Utilizando a técnica de tabulação cruzada, os dados de referência foram combinados com o resultado da classificação. Foram extraídas porcentagens de acertos globais (proporção de pontos dispostos na diagonal principal da matriz) e individuais, bem como os respectivos erros associados com o objetivo de avaliar as causas das confusões. Também foi utilizado o índice kappa como referência de qualidade do mapeamento. De acordo com Congalton e Green (1993), o índice kappa é recomendado como uma medida aceitável de precisão, sendo mais confiável que apenas a média da exatidão global, pois utiliza em um único índice todos os valores da matriz de confusão. Detalhes de como são cálculos os índices mencionados são descritos com maior detalhe em Congalton (1991).

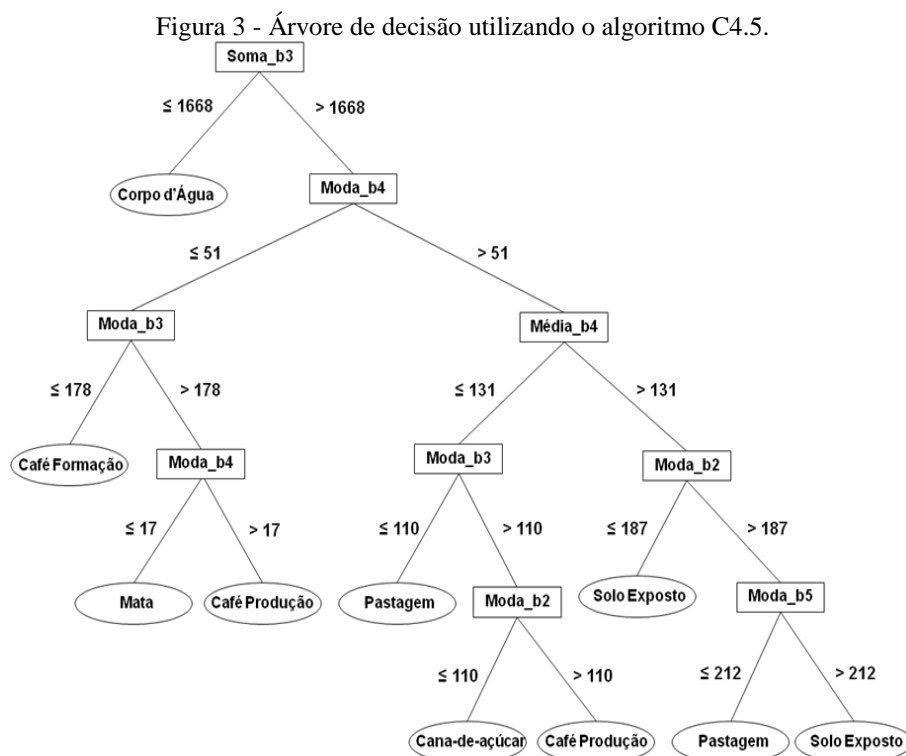
4. RESULTADOS E DISCUSSÃO

Após vários testes de segmentação, adotaram-se como parâmetro de segmentação os seguintes valores: 40 (compacidade), 30 (escala) e 70 (cor), gerando no total 2499 objetos. Como regra geral, os melhores valores são aqueles que conseguem delimitar com eficiência os objetos contidos na imagem sem que haja falta ou excesso de segmentos. A Figura 2 ilustra alguns exemplos de resultados utilizando diferentes valores de entrada.

Figura 2 - Resultados da segmentação utilizando diferentes parâmetros de compacidade/ escala/ cor. Observa-se em (A) excesso de segmentos, em (B) falta de segmentos e em (C) um particionamento da paisagem mais equilibrado.



Com relação à etapa de mineração de dados, a árvore que apresentou o melhor resultado é ilustrada na Figura 3. Conforme pode ser observado, os atributos utilizados na classificação foram a soma, a média e principalmente a moda das propriedades espectrais. Nenhum atributo espacial ou topográfico foi selecionado na construção da árvore de decisão apresentada, o que sugere pouca influência destes atributos no processo de classificação.



A primeira divisão separou a classe água das demais utilizando a soma da banda 3 (comprimento de onda do verde). Os objetos com valores iguais ou abaixo de 1668 foram classificados como Corpo d'água e os objetos com valores acima foram subdivididos.

Na divisão seguinte, objetos com valores de moda na banda 4 (canal vermelho) iguais ou abaixo de 51 e com valores de moda na banda 3 iguais ou inferiores a 178 foram classificados como Café formação. Os objetos com valores acima de 178 e com valores iguais ou inferiores a 17 na moda da banda 4 (infravermelho próximo) foram classificados como Mata. Já os valores maiores que 17 foram então classificados como Café Produção.

Os objetos com valores acima de 51 na moda banda 4 foram subdivididos utilizando a média de seus valores. Objetos com valores iguais ou inferiores a 131 e que apresentaram ao mesmo tempo valores iguais ou inferiores a 110 na moda da banda 3 foram rotulados como Pastagem. Já os objetos com valores acima de 110 e com valores de moda da banda 2 iguais ou inferiores a 110 foram classificados como Cana-de-açúcar. Os valores acima de 110 foram novamente classificados como Café produção.

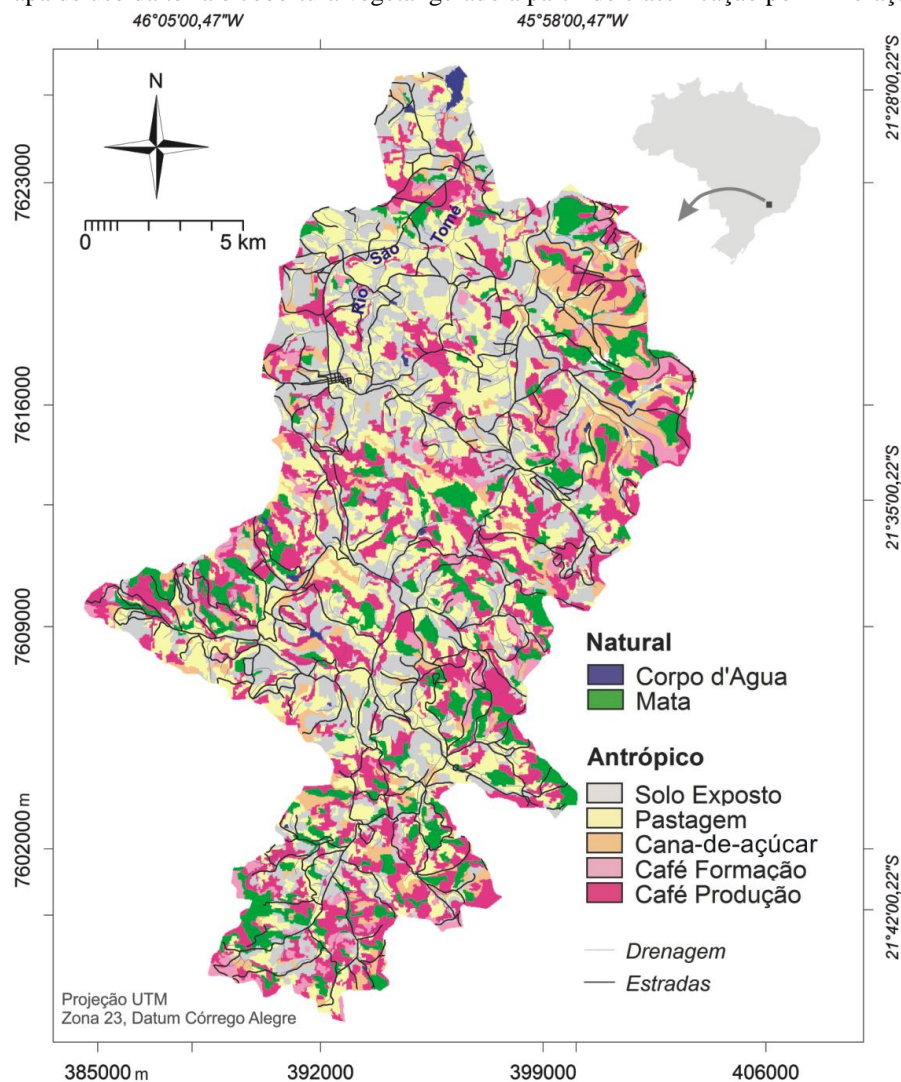
Objetos com valores acima de 131 na média da banda 4 e com moda na banda 2 iguais ou inferiores a 187 foram classificados como Solo exposto. Os objetos com valores acima de 187 e com valores iguais ou inferiores a 212 na moda da banda 5 (infravermelho médio) foram rotulados como Pastagem e os valores acima de 212 como Solo exposto.

Conforme pode ser observado, as classes Pastagem, Solo exposto e Café produção foram classificados em nós distintos da árvore de decisão. Isto aconteceu provavelmente em virtude da heterogeneidade destas classes. A classe Pastagem inclui tanto a vegetação rasteira com baixa atividade fotossintética (pastos secos) quanto os pastos sujos e banhados. A classe Solo exposto também apresenta grande diversidade espectral, pois inclui os solos secos e os solos preparados para o cultivo, além dos afloramentos rochosos. Já a classe Café produção engloba tanto o café em plena atividade produtiva (caracterizado por índices mais altos de área foliar e de sombra de dossel) quanto o café velho ou podado (caracterizado pela maior presença de galhos secos e menor abundância de folhas verdes e de sombra).

Durante os vários testes realizados, algumas árvores selecionaram atributos espaciais. Foram selecionados o perímetro, área, área-caixa, largura e comprimento, todavia, nenhuma delas selecionou atributos topográficos. As classes Solo exposto e Pastagem foram as que mais frequentemente selecionaram os atributos espaciais mencionados, porém, em todas as classificações observou-se superestimativa das classes mencionadas.

O resultado da classificação (Figura 4) ilustra a diversidade de uso e fragmentação espacial da cobertura vegetal presente na bacia. As matas ciliares são praticamente inexistentes e as pequenas manchas localizam-se principalmente nas áreas de encostas. A matriz da bacia é composta por Pastagem, todavia, observa-se uma grande quantidade de Solo exposto em razão da classe também englobar os afloramentos rochosos e haver confusão com os pastos secos. Os plantios de Cana-de-açúcar concentram-se na porção nordeste da bacia por se tratar de uma área mais plana que favorece o cultivo mecanizado. O café, classificado em Formação e Produção é a principal cultura agrícola da bacia. Essa cultura ocupa principalmente as cotas mais altas do relevo, sendo cultivada em vertentes mais íngremes adjacentes às manchas de mata e em topos de morro.

Figura 4 - Mapa de uso da terra e cobertura vegetal gerado a partir de classificação por Mineração de Dados.



Analisando a matriz de confusão (Tabela 1) observa-se que mesmo com uma grande variabilidade de uso existente, os resultados de exatidão (proporção de acertos contidos ao longo da linha) e de confiabilidade (proporção de acertos contidos ao longo da coluna) alcançaram resultados expressivos. A média da exatidão foi de 75% e a média da confiabilidade foi de 68%. A exatidão global que expressa a proporção de pontos corretamente classificados com relação ao total de pontos amostrados alcançou um valor de 73%. Outro índice utilizado como referência para a avaliação da classificação foi o índice Kappa, que alcançou o valor de 0,66, sendo classificada na categoria de desempenho bom (0,61-0,80) (LANDIS; KOCH, 1977).

Tabela 1 - Matriz de confusão gerada para a classificação por mineração de dados no GeoDMA. A média da acurácia (ACC%) foi de 75%. Já a média da confiabilidade em porcentagem foi de 68% e a média global da diagonal principal (exatidão global) foi de 73%.

Classificação → Referências ↓	Café formação	Café produção	Cana-de-açúcar	Pastagem	Solo exposto	Mata	Corpo d'água	Total	ACC%
Café formação	57	13	2	0	0	7	0	79	72%
Café produção	8	219	48	37	7	35	0	354	61%
Cana-de-açúcar	0	0	34	2	1	0	0	37	91%
Pastagem	0	5	1	106	38	2	0	152	69%
Solo exposto	0	1	2	29	188	0	0	220	85%
Mata	2	13	1	2	0	108	0	126	85%
Corpo d'água	0	3	0	1	0	3	12	19	63%
Total	67	254	88	177	234	155	12	987	
Confiabilidade %	85%	86%	38%	59%	80%	69%	100%		

As classes que tiveram as maiores exatidões individuais foram a Cana-de-açúcar (91%) seguida pelo Solo exposto (85%) e Mata (85%). No caso do café, vários estudos já realizados mostram a dificuldade em separar cultivos de café de outros usos em decorrência de sua alta diversidade espectral (MOREIRA et al. 2004, 2008; PRADO et al., 2013). Dependendo do estágio fenológico do cultivo, idade, manejo e condições topográficas, o comportamento espectral do café pode ser semelhante à mata (café produção), ao pasto verde ou cana-de-açúcar (café formação ou café produção em terrenos declivosos) ou pastagem (café velho ou podado). A exatidão da classe Pastagem (69%) foi comprometida pela grande confusão com o Solo exposto. Por fim, a classe Corpo d'água (63%) foi confundida com as áreas de mata e de café produção em virtude provavelmente do sombreamento existentes nestas classes. Com relação à confiabilidade individual, a classe Corpo d'água foi a que apresentou a maior confiabilidade (100%), seguido pelo Café produção (86%) e Café formação (85%). A classe de menor confiabilidade foi a Cana-de-açúcar (38%) em decorrência principalmente da confusão com o Café produção.

5. CONCLUSÃO

A técnica de mineração de dados mostrou-se eficiente no mapeamento do uso da terra e cobertura vegetal em área com grande diversidade de uso como no sul de Minas Gerais. A exatidão global alcançada foi de 73% com índice kappa de 0,66, o que a enquadra na categoria de desempenho bom. Apesar de considerar na mineração de dados atributos espectrais, espaciais e topográficos, apenas informações espectrais foram selecionados na classificação, o que sugere pouca relevância de outros atributos no processo de decisão. Todas as bandas do sensor LISS III (bandas do verde, vermelho, infravermelho próximo e infravermelho médio) foram selecionadas no processo de mineração de dados. Tal resultado reforça o potencial das imagens LISS-III na identificação e classificação do uso da terra e cobertura vegetal.

AGRADECIMENTOS

Os autores expressam os seus agradecimentos a CAPES pelo apoio financeiro concedido. Agradecimentos especiais ao Dr. Thales Korting pelos esclarecimentos com relação ao uso do aplicativo GeoDMA.

REFERÊNCIAS BIBLIOGRÁFICAS

- AB'SÁBER, A. N. Os domínios de natureza no Brasil: potencialidades paisagísticas. 2ª ed. São Paulo: Ateliê Editorial, 2003.
- BAATZ, M. et al. USER GUIDE 4 (eCognition). Definiens Imaging. Copyright © 2000 - 2004. Germany, 2004.
- BAATZ, M.; SCHÄPE, A. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Journal of Photogrammetry and Remote Sensing*, Heidelberg, v. 58, n. 3-4, p. 12-23, 2000.
- BINS, L. S. et al. Satellite Imagery Segmentation: a region growing approach. In: Anais VIII Simpósio Brasileiro de Sensoriamento Remoto, Salvador, Brasil, 14-19, abril, 1996, INPE, p. 677- 680.
- BLASCHKE, T. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2-16, 2010.
- CERON, A. O.; DINIZ, J. A. F. O uso de fotografias aéreas na identificação das formas de utilização agrícola da terra. *Revista Brasileira de Geografia*. n. 2, p.161-173, 1966.
- CHUBEY, M. S. FRANKLIN S. E.; WULDER M. A. Object-based Analysis of Ikonos-2 Imagery for Extraction of Forest Inventory Parameters. *Photogrammetric Engineering & Remote Sensing*. Vol. 72, No. 4, April 2006, pp. 383-394.
- CONGALTON, R. G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *REMOTE SENS. ENVIRON.* 37:35-46 (1991).
- CONGALTON, R. G.; GREEN, K. A Practical Look at the Sources of Confusion in Error Matrix Generation.

- Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 5, May 1993, pp. 641-644.
- CRÓSTA, A. P. Processamento Digital de Imagens de Sensoriamento Remoto. Ed. rev. 4ª Reimpressão. Campinas, SP: IG/UNICAMP, 2002. 170 p.
- DPI/INPE. GeoDMA-*Geographic Data Mining Analyst*. Disponível em: <<http://www.dpi.inpe.br/menu/Projetos/geodma.php>>. Acesso em: 20 de jul. de 2013.
- DO PRADO, D. F. C. et al. Mapeamento do cultivo de café no sul de Minas Gerais utilizando imagens Landsat-5 TM. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto*, Foz do Iguaçu, Brasil, 13-18 de abril 2013, INPE, p. 186-192.
- EMBRAPA – Monitoramento por Satélite. IRS - Indian Remote Sensing Satellite. Disponível em: <http://www.sat.cnpm.embrapa.br/conteudo/missao_irs.php>. Acesso em: 05 de fev. de 2013.
- FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399–408, 1997.
- KAWAKUBO, F. S. et al. Land-use and vegetation-cover mapping of an indigenous land area in the state of Mato Grosso (Brazil) based on spectral linear mixing model, segmentation and region classification. In: Geocarto International. Vol. 24, nº 2, p. 165–175, 2009.
- KAWAKUBO, F. S.; MORATO, R. G.; LUCHIARI, A. Use of fraction imagery, segmentation and masking techniques to classify land-use and land-cover types in the Brazilian Amazon. In: *International Journal of Remote Sensing*. Vol. 34, nº 15, p. 5452–5467, 2013.
- KORTING, T. S. et al. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, Brasil, 25-30 abril 2009, INPE, p. 7813-7820.
- KORTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA – Geographic Data Mining Analyst. In: *Computers & Geosciences*, 57 (2013), 133-145.
- LANDIS, J. R.; KOCH G. G. The Measurement of Observer Agreement for Categorical Data.” *Biometrics* 33: 159–174, 1977.
- MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat-5. *Pesquisa Agropecuária Brasileira*, Brasília, v.39. n.3, p.223-231. 2004.
- MOREIRA, M. A. et al. Geotecnologias no mapeamento da Cultura do Café em Escala Municipal. *Sociedade & Natureza*, Uberlândia, 20 (1): 101-110, jun. 2008.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A.. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11, 1633–1644, 2007.
- QUINLAN, J. R. C4. 5: Programs for Machine Learning. [S.l.]: Morgan kaufmann, 1993.
- RADAMBRASIL, Levantamento dos recursos naturais: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Folhas SF. 23 / 24, Rio de Janeiro/Vitória. Rio de Janeiro: Ministério das Minas e Energia, 1983. 780p.
- SILVA, A. C. et al. Desenvolvimento sustentável da bacia hidrográfica do rio São Tomé. Gestão de Extensão e Estudos Independentes: *Revista do I Fórum de Extensão Universitária*. 2011. 9p.
- SILVA, M. P. S. et al. Remote-sensing image mining: Detecting agents of land-use change in tropical forest areas. *International Journal of Remote Sensing*, 29(16), 4803–4822, 2008.
- STRAHLER, A. N. Hypsometric (area-altitude) – analysis of erosion al topography. *Geological Society of America Bulletin*, v.63, n.10, p.1117-1142, 1952.
- TIAN, J.; CHEN, D. M. Optimization in multi-scale segmentation of high resolution satellite images for artificial feature recognition. *International Journal of Remote Sensing*, Vol. 28, Nº 20, p. 4625–4644, 2007.
- VIEIRA, M. A. et al. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sensing of Environment* 123 (2012) 553-562.