sid.inpe.br/mtc-m19/2013/12.04.03.56-TDI

ÁRVORES DE DECISÃO APLICADAS À EXTRAÇÃO AUTOMÁTICA DE REDES DE DRENAGEM

Lise Christine Banon

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Rafael Duarte Coelho dos Santos, e Nandamudi Lakalapalli Vijaykumar aprovada em 18 de dezembro de 2013.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3FB9RFE>

INPE
São José dos Campos
2013

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
Fax: (012) 3208-6919
E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (RE/DIR-204):
Presidente:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Membros:
Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)
Drª Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)
Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)
Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)
Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Drª Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)
BIBLIOTECA DIGITAL:
Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)
REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)
EDITORAÇÃO ELETRÔNICA:
Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)
André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)

sid.inpe.br/mtc-m19/2013/12.04.03.56-TDI

ÁRVORES DE DECISÃO APLICADAS À EXTRAÇÃO AUTOMÁTICA DE REDES DE DRENAGEM

Lise Christine Banon

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Rafael Duarte Coelho dos Santos, e Nandamudi Lakalapalli Vijaykumar aprovada em 18 de dezembro de 2013.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3FB9RFE>

INPE
São José dos Campos
2013

Dados Internacionais de Catalogação na Publicação (CIP)

Banon, Lise Christine.

B227a Árvores de decisão aplicadas à extração automática de redes de drenagem / Lise Christine Banon. – São José dos Campos : INPE, 2013.

xxiv + 87 p. ; (sid.inpe.br/mtc-m19/2013/12.04.03.56-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.

Orientadores : Drs. Rafael Duarte Coelho dos Santos, e Nandamudi Lakalapalli Vijaykumar.

1. árvore de decisão. 2. mineração de dados. 3. extração. 4. redes de drenagem. 5. MDE SRTM. I.Título.

CDU 004.62:556.51

Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de ***Mestre*** em

Computação Aplicada

Dr. Sergio Rosim

Presidente / INPE / São José dos Campos - SP

Dr. Rafael Duarte Coelho dos Santos

Orientador(a) / INPE / SJCampos - SP

Dr. Nandamudi Lankalapalli Vijaykumar

Orientador(a) / INPE / SJCampos - SP

Dra. Graziela Balda Scofield

Convidado(a) / CEMADEN / Cachoeira Paulista - SP

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

Aluno (a): ***Lise Christine Banon***

São José dos Campos, 18 de Dezembro de 2013

“Jamais considerem seu estudo como uma obrigação, mas sim como uma oportunidade invejável de aprender, sobre a influência libertadora da beleza no domínio do espírito, para seu prazer pessoal e para o proveito da comunidade à qual pertencerá o seu trabalho futuro.”

Albert Einstein *(1879-1955)*

Dedico este trabalho com todo meu amor a meu marido Camilo,
a meus pais Gerald e Marta, a minha querida irmã Gabriela e
as minhas avós Lêda e Michelle (in memoriam)

AGRADECIMENTOS

Agradeço a Deus pela oportunidade de realizar este trabalho.

Ao Dr. Rafael e Dr. Vijay pela orientação e incentivo.

Aos meus professores pelos ensinamentos.

A minha família e a meu marido pelo apoio e carinho durante esta caminhada.

À Coordenação de Aperfeiçoamento de Pessoal de Estudo Superior (CAPES) pela concessão da bolsa de estudo.

RESUMO

A representação cartográfica da Amazônia Legal é uma questão estratégica para o país e sempre representou um grande desafio. A sua extensão territorial ultrapassa 5 milhões de quilômetros quadrados, dos quais cerca de 1,8 milhões não possuem informações cartográficas adequadas, o que tornou esta área conhecida como “vazio cartográfico”. O Projeto Radiografia da Amazônia do Exército Brasileiro tem por objetivo suprimir esta deficiência e conta com a cooperação científica do Instituto Nacional de Pesquisas Espaciais (INPE) para as aplicações dos radares interferométricos que permitem, a partir de seus produtos, a extração de feições cartográficas. A hidrografia é fundamental em uma representação cartográfica. A sua extração é feita a partir de um dos produtos derivados do radar – os modelos digitais de elevação (MDE). Atualmente, a partir destes modelos, estas redes são extraídas automaticamente utilizando a metodologia clássica. Esta metodologia se baseia na escolha empírica de um valor que determina a área de contribuição mínima necessária para a definição da drenagem. No entanto, a região amazônica possui diferentes padrões geomorfológicos e neste caso, os resultados não são satisfatórios. A definição de uma nova metodologia que forneça um resultado mais realista para a extração automática de redes de drenagem traria maior produtividade e, em projetos que envolvem grandes extensões territoriais como o vazio cartográfico, representaria uma contribuição relevante. Técnicas de Mineração de Dados (MD) permitem identificar padrões de interesse em grandes volumes de dados, o que motivou a realização deste trabalho. A hipótese é que a aplicação de árvores de decisão, uma das técnicas de MD, possa indicar critérios e outros atributos extraídos do MDE, que permitam propor uma nova metodologia para a extração automática de uma rede de drenagem mais realista, inclusive em regiões com diferentes padrões geomorfológicos. Neste trabalho, esta técnica foi testada sobre os dados do MDE SRTM (*Shuttle Radar Topography Mission*) para uma área da região amazônica com grande diversidade de padrões geomorfológicos. A metodologia proposta apresentou resultados muito superiores em relação à metodologia clássica. A árvore de decisão obtida indicou que outros atributos, além da área de contribuição, são importantes para a definição de uma rede de drenagem mais realista, minimizando o processo de edição manual. Esta árvore de decisão, resultante da primeira área de estudo, foi aplicada em outras 4 áreas também escolhidas na região amazônica. Em todos os casos, os resultados foram bastante satisfatórios quando comparados aos da metodologia clássica. Desta forma, a árvore de decisão obtida neste trabalho pode ser extrapolável para outras áreas da região amazônica, utilizando os dados do MDE SRTM.

DECISION TREES APPLIED TO THE AUTOMATIC EXTRACTION OF DRAINAGE NETWORKS

ABSTRACT

Legal Amazon region cartographic representation is strategic to Brazil and has always represented a great challenge. Legal Amazon region has a territorial extension of over 5 million square kilometers of which about 1.8 million has inadequate cartographic information and is usually called the "Cartographic Blank of the Amazon". The Amazon Radiography Project developed by Brazilian Army aims to eliminate this deficiency through the scientific cooperation with National Institute for Space Research (INPE) for interferometric radar applications, whose products enable the extraction of cartographic features. The hydrography is essential in a cartographic representation and its extraction is done from a product derived from radar – the digital elevation models (DEM). Currently the drainage networks are automatically extracted from these models, using the classical methodology. This methodology is based on empirical choice of a value that determines the minimum contribution area required to define the drainage. However, the amazon region has different geomorphological patterns and therefore the results are not satisfactory. A new method for automatic extraction of drainage networks could have a more realistic result, providing greater productivity and would mean an important contribution to projects on large territorial extension, as the "Cartographic Blank of the Amazon". Data Mining (DM) techniques allow the identification of patterns of interest in large volumes of data. The hypothesis is that the Decision Trees techniques is able to suggest criteria and other attributes extracted from DEM, leading to a new methodology for automatic extraction of a drainage network more realistic, even in regions with different geomorphological patterns. In this work, this technique was tested on data from DEM SRTM (*Shuttle Radar Topography Mission*) to an area of the amazon region with great diversity of geomorphological patterns. The proposed methodology showed superior results compared to classical methodology. The Decision Tree obtained pointed out that other attributes, besides the contribution area, are important for the definition of a network of more realistic drainage, minimizing the manual editing process. This Decision Tree was applied in 4 other areas also chosen in the amazon region. In all cases the results were very satisfactory compared to the classical methodology. Thus, the Decision Tree obtained in this work can be extrapolated to other areas of the amazon region, using data from the DEM SRTM.

LISTA DE FIGURAS

Pág.

Figura 2.1 – Pré-processamento do MDE no sistema TerraHidro. 7

Figura 2.2 – Exemplo do cálculo da área de contribuição: (a) Identificação das células em uma grade 4x4 (b) Grade altimétrica do MDE com valores em metros, (c) Grade LDD gerada a partir do MDE, (d) Área de Contribuição resultante do cálculo baseado no LDD e (e) Cálculo da Área de Contribuição..... 8

Figura 2.3 – Grade da área de contribuição calculada a partir do LDD (a), destacando-se em azul a rede de drenagem, definida a partir da escolha do limiar da área de contribuição igual a 40 (b) e igual a 10 (c). 11

Figura 2.4 – Área de contribuição (à esquerda) e redes de drenagem representadas (em azul) sobre o MDE SRTM, variando-se os limiares de área de contribuição (em número de células da grade)..... 12

Figura 2.5 – MDE SRTM com a rede de drenagem (em azul) resultante da classificação considerando área de contribuição igual ou superior a: (a) 1.287.900m^2 e (b) 85.860m^2. 13

Figura 2.6 – Processo KDD. 15

Figura 2.7 – Separação perfeita de duas classes no espaço de atributos através da combinação de partições ortogonais (a) usando a árvore de decisão (b). 17

Figura 3.1 – Estudos de caso selecionados na região amazônica, sendo uma área escolhida para a aplicação da metodologia proposta (Área A) e quatro áreas para a elaboração dos testes (Áreas B, C, D e E). 20

Figura 3.2 – Fluxograma geral da metodologia. 22

Figura 3.3 – Exemplos de pontos amostrais representativos das três classes em estudo sobre o MDE SRTM..... 25

Figura 3.4 – Fluxo de atividades: a) Convencional e b) Proposto. 28

Figura 3.5 – Definição de regras booleanas a partir de uma árvore de decisão. 29

Figura 3.6 – Rede de drenagem (em azul) sobre o MDE SRTM: a) antes do pós-processamento, b) indicação da direção de fluxo (em vermelho) e c) resultado do pós-processamento. 29

Figura 3.7 – Rede de drenagem obtida pela metodologia proposta: a) antes do pós-processamento; b) após a correção dos

intervalos com até 5 células; c) após a eliminação de trechos isolados com menos de 3 células (destacados em vermelho em b) e d) conexão de trechos isolados..... 30

Figura 3.8 – Fluxograma do Teste de Simplificação. 35

Figura 3.9 – Fluxograma do Teste de Robustez. 39

Figura 4.1 – Região heterogênea com rede de drenagem (em azul) resultante da metodologia clássica sobre o MDE SRTM. 41

Figura 4.2 – Conjunto amostral resultante da primeira etapa de seleção das amostras após a fase de revisão dos pontos. 42

Figura 4.3 – Árvore de decisão gerada a partir do primeiro conjunto amostral. Para cada folha, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação..... 43

Figura 4.4 – Relação entre da Área de Contribuição e da Ordem Máxima de Sthraler (em valores ranqueados)..... 43

Figura 4.5 – Rede de drenagem (em azul) sobre o MDE SRTM, obtida a partir do primeiro conjunto amostral da metodologia proposta. 44

Figura 4.6 – Conjunto amostral resultante da última etapa de seleção das amostras após a fase de revisão dos pontos. 45

Figura 4.7 – Comparativo entre as árvores geradas pelo J48, usando o Fator de Confiança de 0,25 e 0,1. As folhas F0 a F17 possuem entre parênteses o número total de amostras classificadas e os erros de classificação..... 47

Figura 4.8 – Pontos em vermelho da classe não drenagem, pertencentes as folhas F4 e F5 da Árvore 2, impedem o avanço da rede de drenagem sobre o platô (áreas claras do MDE SRTM)..... 49

Figura 4.9 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) metodologia clássica e b) metodologia proposta. Os círculos destacam as áreas cujos resultados são mais discrepantes..... 51

Figura 4.10 – Exemplos de duas avaliações com a seleção em vermelho do melhor resultado..... 52

Figura 4.11 – Cruzamento das redes de drenagem obtidas a partir da metodologia clássica e proposta sobre o MDE SRTM. Em azul são os trechos comuns às duas redes, em ciano os exclusivos da metodologia proposta e em vermelho os exclusivos da metodologia clássica..... 52

Figura 4.12 – Árvore de decisão 3 utilizando apenas duas classes. Para as folhas F0 a F8, os números entre parênteses

representam, respectivamente, o total de amostras classificadas e os erros de classificação. 54

Figura 4.13 – Cruzamento das redes de drenagem obtidas a partir da utilização de 2 e 3 classes sobre o MDE SRTM. Em azul são representados os trechos comuns às duas redes, em verde os exclusivos para a abordagem usando 3 classes e em vermelho os exclusivos para 2 classes. 55

Figura 4.14 – Exemplos de duas avaliações com o melhor resultado selecionado com um retângulo vermelho. Os círculos destacam as áreas mais discrepantes. 56

Figura 4.15 – Árvore de decisão 4 utilizando apenas atributos morfométricos. Para as folhas F0 a F13, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação. 58

Figura 4.16 – Gráfico indicando o número de erros por folha da Árvore de decisão 4. 59

Figura 4.17 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) abordagem usando apenas atributos morfométricos e b) metodologia proposta. Os círculos destacam algumas áreas divergentes. 60

Figura 4.18 – Exemplos de duas avaliações com o melhor resultado selecionado com um retângulo vermelho. 60

Figura 4.19 – Árvore de decisão obtida pela abordagem usando apenas a área de contribuição. Para cada folha, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação. 61

Figura 4.20 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) abordagem usando apenas a área de contribuição e b) metodologia clássica. Os círculos destacam algumas áreas divergentes. 62

Figura 4.21 – Rede de drenagem sobreposta ao MDE SRTM da área B. a) com as células referentes a folha F10 destacadas em vermelho e a demais células da rede de drenagem em azul e b) rede de drenagem (em azul) resultante da conexão de todos os trechos isolados. 64

Figura 4.22 – MDE SRTM da Área B com área circular destacando uma região antropizada, dificultando o reconhecimento das nascentes. Ao lado, a imagem Geocover indicando esta influência (áreas magentas). 65

Figura 4.23 – Comparação entre os resultados das metodologias clássica e proposta numa quadrícula da área B. a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede

de drenagem obtida pela aplicação da árvore da metodologia proposta usando *gapsize* e *segsize* iguais a zero e d) rede de drenagem obtida pela aplicação da árvore da metodologia proposta usando *gapsize* = 5 e *segsize* = 3. O círculo destaca uma área discrepante. 66

Figura 4.24 – Resultados da Área B: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino..... 68

Figura 4.25 – Resultados da Área C: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino..... 69

Figura 4.26 – Resultados da Área D: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino..... 70

Figura 4.27 – Resultados da Área E: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino..... 71

Figura A.1 – Localização e número das células da subgrade 3x3. 79

Figura A.2 – Exemplo de MDE e LDD 82

Figura A.3 – Exemplo de cálculo do atributo areacontr (em células) 83

Figura A.4 – Exemplo de cálculo do atributo desvtopo 84

Figura A.5 – Exemplo de cálculo do atributo desvvert 85

Figura A.6 – Exemplo de cálculo do atributo decld8 85

Figura A.7 – Exemplo de cálculo do atributo ordstrmax 86

Figura A.8 – Exemplo de cálculo do atributo hand_ordstrmax3. As células marcadas em azul possuem valores de ordstrmax igual ou

superior a 3. As setas em vermelho mostram o caminho que conecta a célula vermelha à verde..... 87

Figura A.9 – Exemplo de cálculo do atributo difz. As setas em vermelho mostram o caminho que conecta a célula vermelha à verde. 87

LISTA DE TABELAS

Pág.

Tabela 3.1 – Características das áreas escolhidas para os quatro estudos de caso..... 19
Tabela 3.2 – Lista de atributos morfométricos..... 23
Tabela 3.3 – Lista de atributos baseados na direção de fluxo..... 24
Tabela 3.4 – Lista de parâmetros do algoritmo J48. 27
Tabela 4.1 – Matriz de confusão obtida na Fase Final da metodologia proposta. 50
Tabela 4.2 – Matriz de confusão obtida pela abordagem com a simplificação do número de classes..... 53
Tabela 4.3 – Matriz de confusão obtida pela abordagem usando apenas atributos morfométricos..... 57
Tabela 4.4 – Matriz de confusão obtida pela abordagem simplificada usando apenas a área de contribuição. 61
Tabela 4.5– Representatividade das folhas pertencentes as classes de nascente e drenagem da árvore de decisão da metodologia proposta na área de treinamento e nas áreas teste. 63
Tabela 4.6– Resultado das 50 Avaliações do Ganho Significativo..... 67

LISTA DE SIGLAS E ABREVIATURAS

DSG	Diretoria de Serviço Geográfico
ENVI	*The Environment for Visualizing Images*
IBGE	Instituto Brasileiro de Geografia e Estatística
IBL	*Instance-Based Learning*
IDL	*Interactive Data Language*
INPE	Instituto Nacional de Pesquisas Espaciais
KDD	*Knowledge Discovery in Databases*
LDD	*Local Drainage Direction*
MD	Mineração de Dados
MDE	Modelo Digital de Elevação
SAR	*Synthetic Aperture Radar*
SRTM	*Shuttle Radar Topography Mission*
SUDAM	Superintendência do Desenvolvimento da Amazônia
WEKA	*Waikato Environment for Knowledge Analysis*

SUMÁRIO

Pág.

1 INTRODUÇÃO 1
1.1. Hipótese 4
1.2. Objetivo 4
2 FUNDAMENTAÇÂO TEÓRICA 5
2.1. Modelo Digital de Elevação 5
2.2. Direção de Fluxo e Área de Contribuição 6
2.3. Metodologia clássica para extração automática de redes de drenagem baseada na área de contribuição 10
2.4. Outros atributos utilizados na extração de redes de drenagem 14
2.5. Mineração de Dados 15
3 MATERIAL E MÉTODOS 19
3.1. Área de estudo 19
3.2. Material 21
3.3. Metodologia 21
3.3.1. Pré-processamento e Extração de Atributos 23
3.3.2. Metodologia clássica 24
3.3.3. Metodologia proposta 24
3.3.3.1. Seleção inicial das amostras e verificação 25
3.3.3.2. Seleção de atributos, classificação e pós-processamento 26
3.3.3.3. Avaliação da rede de drenagem e seleção de novas amostras 31
3.3.4. Comparação entre as redes de drenagens obtidas pelas metodologias clássica e proposta 34
3.3.5. Testes de Simplificação 34
3.3.5.1. Simplificação do número de classes 36
3.3.5.2. Simplificação com a utilização exclusiva dos atributos morfométricos 36
3.3.5.3. Simplificação através da definição do limiar ótimo para a área de contribuição 37
3.3.6. Testes de Robustez 38
4 RESULTADOS E DISCUSSÃO 41
4.1. Rede de drenagem obtida a partir da metodologia clássica 41
4.2. Rede de drenagem obtida a partir da metodologia proposta 42

4.2.1. Fase Inicial: Seleção das Amostras, Árvore de Decisão e Avaliação 42
4.2.2. Fase Final: Seleção das Amostras, Árvore de Decisão e Avaliação 45
4.2.3. Testes de Simplificação 53
4.2.3.1. Simplificação do número de classes 53
4.2.3.2. Simplificação usando apenas atributos morfométricos 56
4.2.3.3. Simplificação usando apenas a área de contribuição 61
4.2.4. Testes de Robustez 63
5 CONSIDERAÇÕES FINAIS 73
REFERÊNCIAS BIBLIOGRÁFICAS 75
APÊNDICE A –CÁLCULO DOS ATRIBUTOS 79
A.1 - Atributos Morfométricos – método de Evans-Young 79
A.2 - Atributos baseados na direção de fluxo 82

1 INTRODUÇÃO

A Amazônia Legal é uma divisão territorial político-administrativa definida em 1966, com a criação da Superintendência do Desenvolvimento da Amazônia (SUDAM). Compreende o bioma Amazônia e 20% do bioma Cerrado e é constituída por nove estados brasileiros (Acre, Amapá, Amazonas, parte oeste do Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins). A Amazônia Legal possui uma extensão superior a 5 milhões de quilômetros quadrados e contém 1/3 das florestas tropicais úmidas mundiais, sendo uma região de grande importância econômica, científica e socioambiental (IBGE, 2012).

Devido a sua relevância, a representação cartográfica da região amazônica é uma questão estratégica para o país e sempre representou um grande desafio. Além da grande dimensão territorial e da densa floresta que recobre a região amazônica, há uma constante presença de nuvens e adversidades climáticas. Estes fatores dificultam e inviabilizam métodos tradicionais de recobrimento aerofotogramétrico e imageamento óptico por satélites, que são necessários para a cartografia desta região (DSG, 2008).

Cerca de 1,8 milhões de quilômetros quadrados da Amazônia Legal não possuem informações cartográficas terrestres adequadas em escalas superiores a 1:250.000, o que tornou esta área conhecida como região do “vazio cartográfico”. Além disso, o mapeamento existente nesta região é falho, sendo composto por documentos cartográficos que representam o nível da copa das árvores e não as feições planialtimétricas no nível do solo (CORREIA, 2011).

A Diretoria de Serviço Geográfico (DSG) é o órgão responsável pela direção e fiscalização das atividades cartográficas no âmbito do Exército Brasileiro e dos convênios estabelecidos entre o Exército e os órgãos públicos. A partir do ano 2000, a DSG e o Instituto Nacional de Pesquisas Espaciais (INPE) iniciaram a experimentação da tecnologia de radares SAR (*Synthetic Aperture Radar*)

interferométricos nas bandas “X” e “P” para aplicações cartográficas em regiões de floresta tropical densa. Enquanto a radiação na banda “X” permite a interação apenas no nível da copa das árvores, as ondas eletromagnéticas na banda “P” penetram no dossel da floresta e interagem com a superfície do terreno. Os bons resultados apresentados com o uso desta tecnologia (DUTRA et al., 2002) motivaram a DSG a iniciar em 2008 o projeto intitulado Radiografia da Amazônia (CORREIA, 2011; MOURA; CORREIA, 2011). O objetivo deste projeto é suprimir o atual vazio cartográfico, a partir da obtenção de diversos produtos, dentre os quais constam os modelos digitais de elevação – MDE.

A hidrografia é uma importante feição cartográfica que, tradicionalmente, pode ser extraída a partir de dois métodos: a estereoscopia com o uso de ortoimagens ou a extração automática a partir dos MDE. O primeiro método implica em um considerável custo financeiro e demanda tempo para ser executado, inviabilizando-o como metodologia para uma região de grandes extensões. O segundo método, adotado atualmente pelo Projeto Radiografia da Amazônia (CORREIA, 2011), utiliza a metodologia clássica para a extração automática das redes de drenagem, ou seja, baseia-se exclusivamente na escolha empírica de um valor que determina a área de contribuição mínima necessária para a definição da drenagem (O'CALLAGHAN; MARK, 1984). A área de contribuição, ou área acumulada, é um atributo extraído do MDE. O valor da área de contribuição de uma célula da grade corresponde ao número de células que convergem para esta célula em questão, considerando a direção de fluxo.

A aplicação da metodologia clássica em regiões da Amazônia não apresentou bons resultados (PINHEIRO et al., 2009). A região amazônica compreende áreas com diferentes padrões geomorfológicos, e neste caso, basear a extração de feições hidrológicas a partir do limiar de um único atributo, torna-se insuficiente para a obtenção de uma drenagem representativa, demandando um árduo trabalho de edições manuais, como a eliminação de feições inexistentes e a inclusão de trechos não detectados.

A partir do MDE é possível gerar uma grande quantidade de atributos, além da área de contribuição, como por exemplo, os atributos do terreno (VALERIANO et al., 2006). Destes atributos, alguns estão associados à hidrologia, tais como as curvaturas vertical e horizontal que podem indicar áreas de convergência ou divergência dos fluxos de água na superfície, auxiliando na identificação de regiões potencialmente associadas à rede de drenagem. Desta forma, a utilização de outros atributos poderia contribuir para a obtenção de uma rede mais realista. Seguindo este conceito, a proposta de uma nova metodologia utilizando vários atributos do MDE poderia ser definida para a extração automática de redes de drenagem, a fim de obter uma maior qualidade do resultado, principalmente em regiões com diferentes padrões geomorfológicos.

A obtenção de uma rede de drenagem mais realista diminui a demanda por edições manuais, aumentando a eficiência do processo de extração das feições hidrológicas. Além disso, as edições manuais seguem critérios subjetivos sendo, portanto, de difícil reprodutibilidade e podem interferir na qualidade final da rede de drenagem. O ganho em eficiência e qualidade tem impacto ainda maior quando os processamentos envolvem grandes extensões territoriais como, por exemplo, no Projeto Radiografia da Amazônia. Sendo assim, este avanço representaria uma contribuição relevante em aplicações cartográficas.

Atualmente, não há um processo completo para a extração automática de redes de drenagem que tenha fornecido um resultado satisfatório para a região amazônica. Novos critérios baseados em vários atributos poderiam ser testados na obtenção de feições hidrológicas mais realistas. Estes critérios e atributos poderiam ser avaliados e selecionados a partir da aplicação da Mineração de Dados (MD) – uma metodologia bastante eficiente para identificar padrões de interesse em grandes volumes de dados.

1.1. Hipótese

A fim de auxiliar o processo de estruturação de uma nova metodologia para a extração automática de redes de drenagem, surge a seguinte questão científica: a partir da aplicação de técnicas de MD, é possível definir critérios baseados em outros atributos (além da área de contribuição) que permitam a extração automática de redes de drenagem de forma mais realista, inclusive em regiões com diferentes padrões geomorfológicos?

1.2. Objetivo

O objetivo principal deste trabalho é propor uma nova metodologia para a extração automática de redes de drenagem capaz de representar áreas com diferentes padrões geomorfológicos. A definição de critérios baseados em um conjunto de atributos mais representativo será obtida através da aplicação do algoritmo J48, um classificador do tipo árvore de decisão, em várias regiões da Amazônia. Com a finalidade de atingir o objetivo descrito acima, foram elaborados os seguintes objetivos específicos:

- Avaliar a qualidade da rede de drenagem extraída automaticamente utilizando a metodologia clássica e a metodologia proposta.

- Comparar a metodologia clássica e a metodologia proposta, avaliando a importância da inclusão de novos atributos para a extração automática da rede de drenagem.

- Avaliar a robustez da árvore de decisão obtida pela metodologia proposta, aplicando-a em áreas distintas da região na qual esta árvore foi obtida.

2 FUNDAMENTAÇÂO TEÓRICA

A metodologia clássica para a extração automática de redes de drenagem é uma técnica que se baseia em um único atributo: a área de contribuição. Este atributo é calculado a partir da grade de direção de fluxo, que por sua vez é extraída a partir do Modelo Digital de Elevação. Este capítulo apresenta uma descrição teórica destes conceitos, definindo e exemplificando os principais elementos que fundamentam esta técnica. Em seguida, são discutidos alguns temas como a utilização de outros atributos para a extração de redes de drenagem e um contexto geral sobre as aplicações da mineração de dados com ênfase nas árvores de decisão.

2.1. Modelo Digital de Elevação

O Modelo Digital de Elevação – MDE, termo derivado do inglês *Digital Elevation Models* – DEM, representa a altimetria de uma região e pode ser estruturado de três formas: grades regulares, triangulares e curvas de nível. Das estruturas de dados, as grades regulares são as mais empregadas para representar um MDE. Devido a sua fácil implementação e por ser compatível com os dados de sensoriamento remoto, são amplamente utilizadas em sistemas de informações geográficas (COLLINS E MOON, 1981).

Os dados da Missão do Ônibus Espacial de Topografia por Radar (*Shuttle Radar Topography Mission* - SRTM) tornaram-se disponíveis para cerca de 80% da superfície terrestre a partir do ano 2000 (VAN ZYL, 2001). O MDE SRTM possui resolução horizontal de 3 arcos de segundos (cerca de 90 m próximo à linha do Equador), resolução vertical de um metro, o que para muitas regiões revela-se como o melhor e mais preciso dado topográfico disponível.

No entanto, devido à geração de seus dados, o MDE SRTM possui algumas limitações e por isso nem sempre representa fielmente os fluxos locais superficiais. Para minimizar este problema, é necessário executar uma fase de pré-processamento dos dados, na qual o MDE é corrigido a partir da eliminação dos sumidouros (mínimos locais). O pré-processamento resulta na

grade de direção de fluxo, que representa os caminhos contínuos da água de um ponto até uma das bordas da grade. Esta correção é imprescindível, pois os atributos baseados na direção de fluxo são fundamentados nas informações contidas nesta grade. Desta forma, na fase de pré-processamento são gerados dois produtos hidrologicamente consistidos: o MDE e a grade de direção de fluxo. Esta etapa de pré-processamento do MDE SRTM pode ser feita utilizando, por exemplo, o Sistema TerraHidro (ROSIM et al., 2003; ABREU et al., 2012).

A Figura 2.1 apresenta um exemplo do pré-processamento realizado pelo TerraHidro para a eliminação de sumidouros (identificados nas grades de direção de fluxo com o símbolo de um quadrado). Note que das cem células presentes na grade do MDE original, dez foram alteradas (células identificadas com a cor de fundo amarela), provocando a mudança de direção de 36 células da grade de direção de fluxo consistida (identificadas pelas setas em vermelho). Também é possível observar que dos nove sumidouros presentes na grade de direção de fluxo original, seis foram eliminados, restando apenas os três localizados na borda da grade de direção de fluxo consistida.

O MDE original difere do MDE hidrologicamente consistido, pois neste último as altimetrias são modificadas para garantir um fluxo contínuo. Esta modificação é feita através de preenchimentos de depressões locais ou pela remoção de obstruções para a abertura de novos caminhos preferenciais dos fluxos de água.

2.2. Direção de Fluxo e Área de Contribuição

A direção de fluxo define a direção preferencial do escoamento da água na superfície e é usualmente identificada pela sigla LDD (derivada do termo em inglês *Local Drainage Direction*). O LDD é extraído do MDE, a partir da identificação da direção de maior declividade.

Indiretamente, a área de contribuição também pode ser considerada um atributo do MDE, pois é calculada a partir do LDD. A área de contribuição

também é conhecida como área acumulada, uma vez que representa a área a montante cujos fluxos convergem para um determinado ponto.

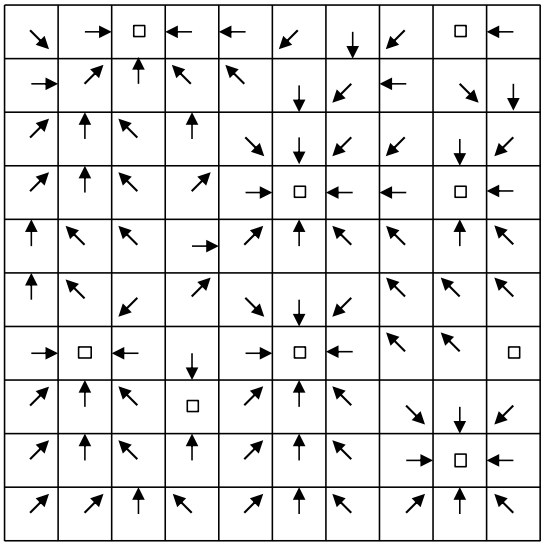
MDE ORIGINAL

42	41	34	39	46	46	45	44	44	45
44	38	41	40	41	44	43	46	46	45
43	39	42	44	41	40	44	45	45	43
40	42	46	47	45	39	40	44	41	42
41	43	44	47	43	46	41	42	45	47
44	46	45	46	46	44	42	44	47	49
45	42	45	46	47	40	45	45	47	47
46	43	46	44	45	43	46	45	47	48
47	46	45	46	46	44	45	45	43	47
48	47	47	47	46	45	46	45	47	47

MDE CONSISTIDO

42	41	34	39	46	46	45	44	44	45
44	38	41	35,67	41	44	43	46	46	45
43	39	42	44	37,33	40	44	45	45	43
40	42	46	47	45	39	39,25	44	41	42
41	43	44	47	43	46	41	39,50	45	47
41,50	46	45	46	46	44	39,75	44	47	49
45	42	45	46	47	40	45	45	47	47
46	43	46	45,75	45	40,75	46	45	47	48
47	46	45	46	46	44	41,50	45	43	47
48	47	47	47	46	45	46	42,25	47	47

GRADE DE DIREÇÃO DE FLUXO ORIGINAL



GRADE DE DIREÇÃO DE FLUXO CONSISTIDO

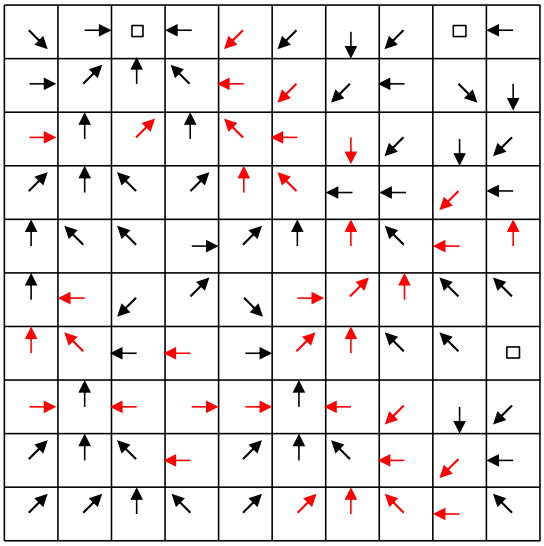


Figura 2.1 – Pré-processamento do MDE no sistema TerraHidro.

A Figura 2.2 apresenta o cálculo da área de contribuição, através de uma grade 4x4 cujas células são identificadas na Figura 2.2a. Uma grade MDE contendo os dados de altimetria é exemplificada na Figura 2.2b. Cada célula da grade MDE representa a altitude média da superfície de um determinado espaço

geográfico. Como o deslocamento da água na superfície sempre percorre no sentido da maior declividade (desnível em relação à distância horizontal), torna-se possível definir o sentido de convergência de cada célula, formando-se uma rede (Figura 2.2c), com a qual será computada a área de contribuição.

a

c11	c12	c13	c14
c21	c22	c23	c24
c31	c32	c33	c34
c41	c42	c43	c44

Identificação das células na grade

b

24	22	22	23
23	19	21	22
22	18	15	21
22	21	12	21

Valores na grade MDE

c

↘	↓	↙	↙
↘	↘	↓	↙
→	↘	↓	↙
↗	→		←

Direção de Fluxo LDD

d

1	1	1	1
1	4	2	1
1	4	8	1
1	1	16	1

Valores na grade Área de Contribuição

e

Cálculo dos valores das células da Área de Contribuição a partir do LDD

[c**] = 1

[c22] = [c22] + [c11] + [c12] + [c13]= 1 + 1 + 1 + 1 = 4

[c23] = [c23] + [c14] = 1 + 1 = 2

[c32] = [c32] + [c21] + [c31] + [c41] = 1 + 1 + 1 + 1 = 4

[c33] = [c33] + [c22] + [c23] + [c24] = 1 + 3 + 3 + 1 =8

[c43] = [c43] + [c42] + [c32] +[c33] + [c34] + [c44] = 1 + 1 + 4 + 8 + 1 + 1 = 16

Figura 2.2 – Exemplo do cálculo da área de contribuição: (a) Identificação das células em uma grade 4x4 (b) Grade altimétrica do MDE com valores em metros, (c) Grade LDD gerada a partir do MDE, (d) Área de Contribuição resultante do cálculo baseado no LDD e (e) Cálculo da Área de Contribuição.

Para o cálculo da área de contribuição, inicialmente todas as células recebem o valor igual a 1. O valor de uma determinada célula será o resultado da soma do número 1 (valor inicial da célula) com os valores das células vizinhas que convergem para esta célula em questão. Por definição, as células que não possuem outras células convergindo em sua direção, permanecem com o valor

da área de contribuição igual a 1. A Figura 2.2e apresenta o cálculo dos valores para cada célula pertencente à grade Área de Contribuição. Por exemplo, a célula c32 desta grade (Figura 2.2d) possui o valor igual a 4. De acordo com a grade LDD (Figura 2.2c), as células c21, c31 e c41 convergem em direção à célula c32. Como as células c21, c31 e c41 não possuem outras que convergem em sua direção, todas são mantidas com o valor inicial igual a 1. Desta forma, o valor da célula c32 será igual à soma dos valores das células c21, c31, c41 e o valor inicial da própria célula c32, que também é 1, totalizando o valor 4. A Figura 2.2d apresenta um exemplo dos valores da área de contribuição calculada para cada célula a partir da grade LDD (Figura 2.2c). O atributo área de contribuição, que neste exemplo está expresso em número de células, também pode ser expresso em área (como por exemplo, m^2 ou km^2). O algoritmo a seguir descreve resumidamente o procedimento para o cálculo dos valores das células da Área de Contribuição, exemplificado na Figura 2.2e:

```
PROCEDURE ACCUMULATION(cell)
   accum(cell) ← 1
   neighbourcell ← NEIGHBOUR(cell);return grid cells draining into the cell
   IF neighbourcell <> NULL THEN
      FOR EACH neighbourcell DO
         IF accum(neighbourcell) = 0 THEN ACCUMULATION(neighbourcell)
         accum(cell) ← accum(cell) + accum(neighbourcell)
      ENDFOR
   ENDIF
END
FOR EACH cell DO
   IF accum(cell) = 0 THEN ACCUMULATION(cell)
ENDFOR
```

2.3. Metodologia clássica para extração automática de redes de drenagem baseada na área de contribuição

Um dos primeiros trabalhos sobre a extração automática de redes de drenagem, a partir do MDE, foi apresentado por O'Callaghan e Mark (1984). Esta metodologia clássica propõe o uso de um limiar baseado na área de contribuição, ou seja, escolhe-se empiricamente um valor mínimo para a área de contribuição. Todas as células da grade acima deste valor são definidas como pontos da rede de drenagem.

Quanto maior o valor acumulado em uma célula, maior a possibilidade de associação desta célula a uma rede de drenagem e, dependendo do limiar escolhido, a célula pode ser incluída nesta rede, conforme exemplificado na Figura 2.3. A partir do LDD (Figura 2.3a), obtém-se a grade da área de contribuição. Neste exemplo, caso o limiar escolhido fosse 40, a rede de drenagem obtida seria formada pelas células marcadas em azul na Figura 2.3b. Em contrapartida, se o limiar fosse reduzido para 10, a rede de drenagem sofreria uma expansão, cujo resultado pode ser observado na Figura 2.3c.

Métodos baseados na escolha de um limiar para a área de contribuição são muito simples de implementar e portanto muito populares. No entanto, em regiões com diferentes padrões geomorfológicos, a escolha de um único limiar para toda a região não é representativo, podendo gerar redes de drenagem com maior ou menor densidade do que a real. Rennó et al. (2008) testaram diferentes limiares em uma região próxima à Manaus/AM, usando o MDE SRTM, cujo resultado pode ser observado na Figura 2.4. Note que quanto menor o limiar, maior a densidade da rede de drenagem.

Numa região da Amazônia Legal, próxima ao município de Santarém/PA, Banon et al. (2013) testaram a metodologia clássica em uma área com diferentes padrões geomorfológicos. A Figura 2.5a apresenta o MDE SRTM desta região. A escolha de um alto limiar para a área de contribuição resultou em uma adequada classificação para as áreas com menor densidade de

drenagem, mas em contrapartida não classificou adequadamente as regiões mais densas. Na área ampliada da região mais densa, é possível observar que as extremidades da drenagem não foram classificadas. Por outro lado, a Figura 5b apresenta a mesma região, desta vez usando um baixo limiar, o que resultou em uma adequada classificação para as áreas com maior densidade de drenagem, mas criou feições inexistentes em áreas com menor densidade. Uma vez que, em regiões com diferentes padrões geomorfológicos, os resultados não são satisfatórios utilizando a metodologia clássica, há a necessidade de edições manuais ou semiautomáticas. Esta etapa pode ser bastante complexa e demorada e, portanto, dispendiosa.

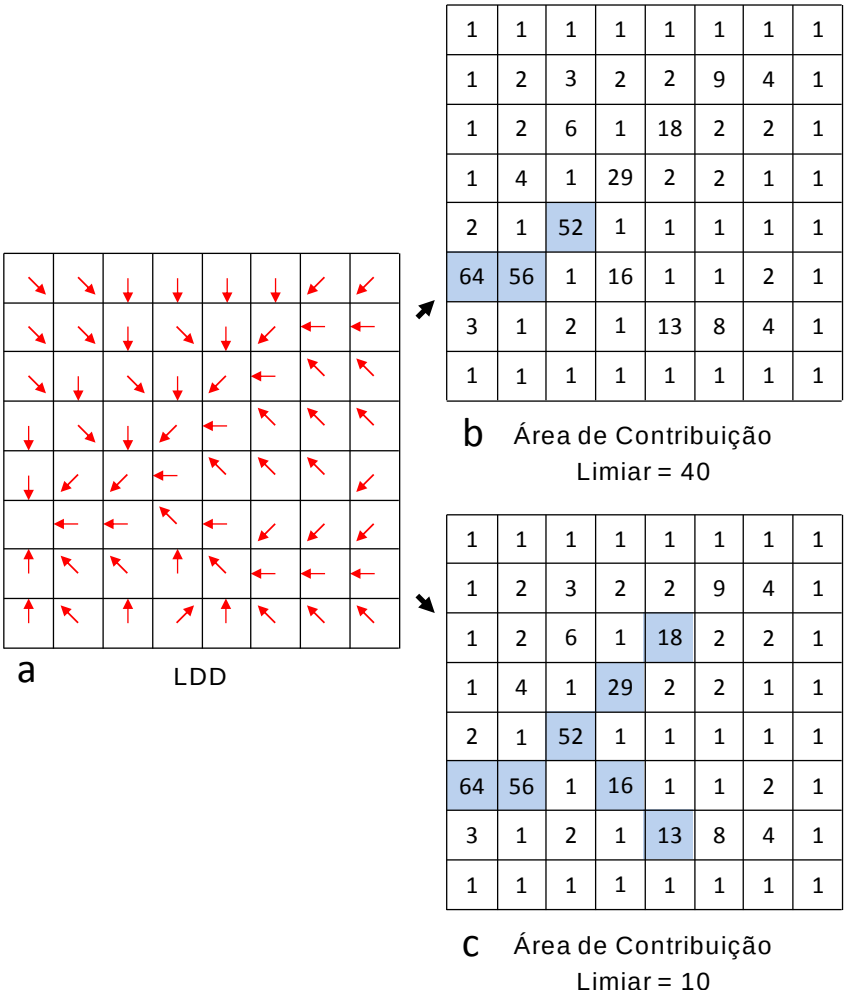


Figura 2.3 – Grade da área de contribuição calculada a partir do LDD (a), destacando-se em azul a rede de drenagem, definida a partir da escolha do limiar da área de contribuição igual a 40 (b) e igual a 10 (c).

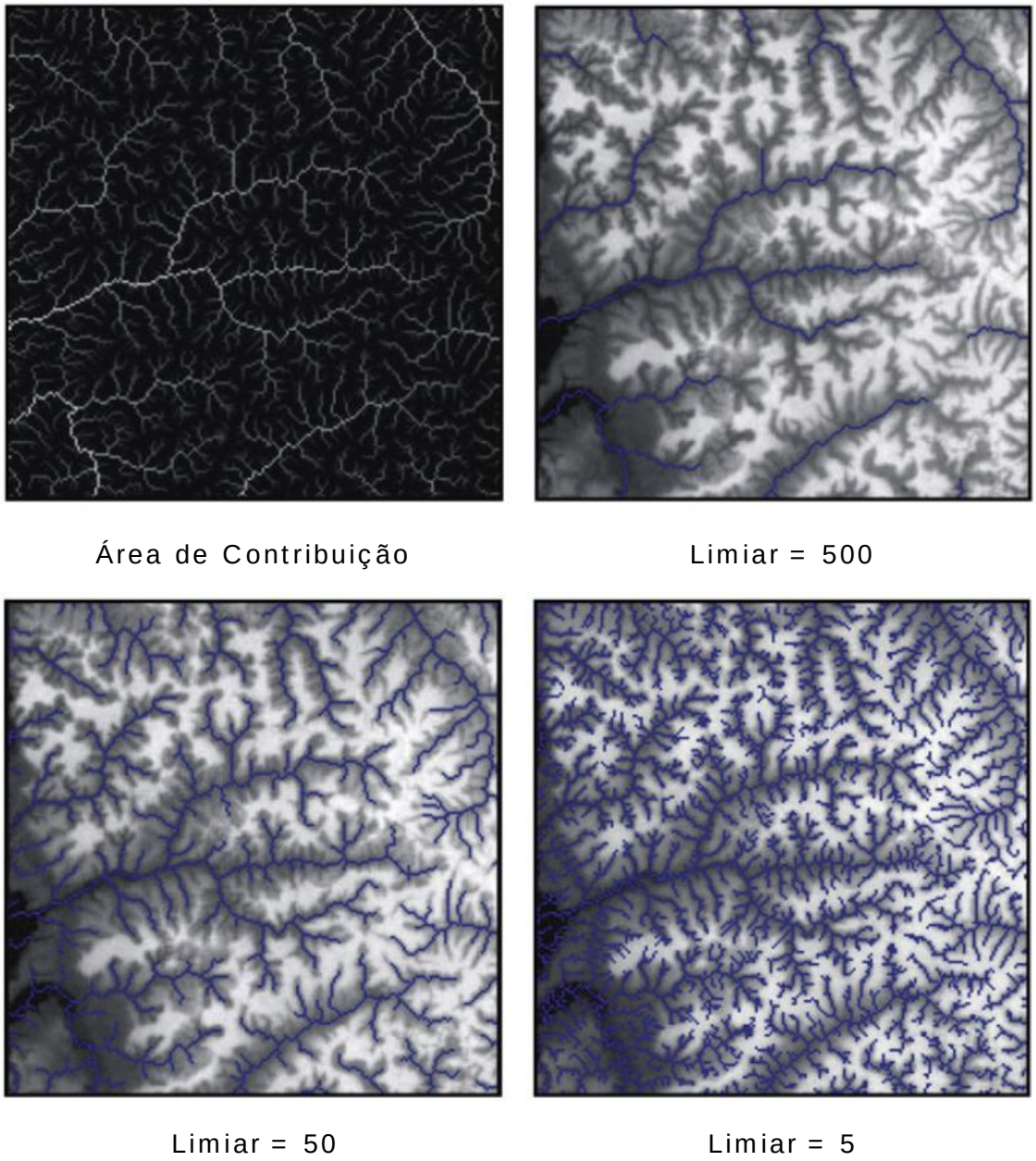


Figura 2.4 – Área de contribuição (à esquerda) e redes de drenagem representadas (em azul) sobre o MDE SRTM, variando-se os limiares de área de contribuição (em número de células da grade)
Fonte: Modificado de Rennó et al.(2008)

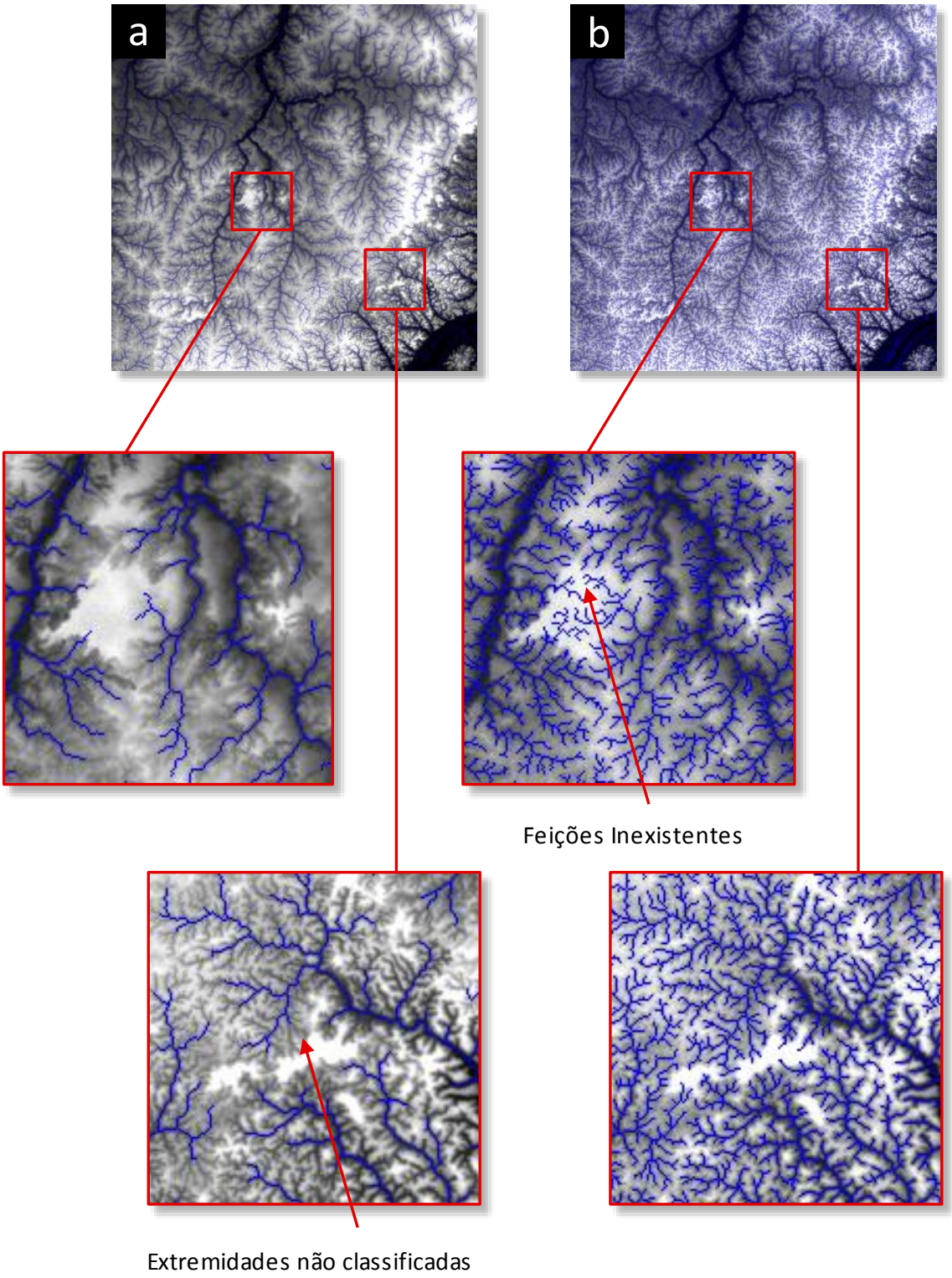


Figura 2.5 – MDE SRTM com a rede de drenagem (em azul) resultante da classificação considerando área de contribuição igual ou superior a: (a) 1.287.900m^2 e (b) 85.860m^2.

Fonte: Banon et al.(2013)

2.4. Outros atributos utilizados na extração de redes de drenagem

A metodologia clássica utiliza apenas a área de contribuição para a identificação dos pontos pertencentes à rede de drenagem. Os critérios utilizados na escolha do limiar para a área de contribuição são muito subjetivos, o que torna a extração da rede de drenagem um processo empírico e dependente do conhecimento e experiência prévia de quem faz essa extração, sendo de difícil reprodutibilidade.

Nos últimos anos, muitos métodos de extração de redes de drenagem têm sido desenvolvidos a partir de uma grade regular utilizando vários atributos do MDE. Esses atributos podem ser agrupados em dois tipos: os que representam as feições geomorfológicas e os baseados em características hidrológicas do terreno (SOILLE et al., 2003). A área de contribuição é um atributo que pertence a este último grupo.

Segundo Tarboton et al. (1991 e 1992), as nascentes são definidas como regiões onde os processos convergentes de fluxos dominam os processos divergentes. Estes processos podem ser representados pela curvatura horizontal, um dos atributos extraídos do DEM. Além disso, usualmente estas nascentes localizam-se em regiões de encosta, onde ocorre a mudança da forma convexa para a côncava. O atributo curvatura vertical descreve estas formas de relevo. Nesta região de transição, ocorre um processo de desaceleração dos fluxos de água e aumento de deposição de materiais, pela diminuição da declividade local.

Desta forma, Tarboton et al. (1991) propõem um critério objetivo para a definição da rede de drenagem baseado na relação entre a declividade e a área de contribuição. Após alguns anos, Tarboton e Ames (2001) sugerem o uso de curvaturas locais extraídas do MDE como uma ponderação para a área de contribuição, o que, segundo os autores, permitiria a extração de redes de drenagem para regiões com diferentes padrões geomorfológicos.

Entretanto, testes realizados por Rennó et al. (2008) mostraram que estas abordagens não obtiveram bons resultados quando aplicadas à região amazônica, utilizando-se o MDE SRTM. Estes autores produziram redes de drenagens mais realistas quando utilizaram, além da área de contribuição, outros dois critérios restritivos baseados em simplificações das curvaturas horizontal e vertical.

2.5. Mineração de Dados

Diante de um volume considerável de dados, é possível definir técnicas que permitam identificar informações relevantes, entender melhor o significado destes dados e estipular procedimentos de tomada de decisão. Para esta finalidade, surgiu a Descoberta do Conhecimento em Banco de Dados, termo derivado do inglês *Knowledge Discovery in Databases* (KDD).

Segundo Fayyad et al. (1996), o KDD (Figura 2.6) pode ser definido como um processo não trivial de identificar em dados, padrões que sejam válidos, novos (agreguem conhecimento), potencialmente úteis e compreensíveis. A Mineração de Dados (MD) é um passo fundamental dentro deste processo maior conhecido como KDD e compreende a seleção dos métodos que serão utilizados para identificar regras e padrões de interesse nos dados em análise, sempre buscando o melhor ajuste dos parâmetros no algoritmo escolhido.

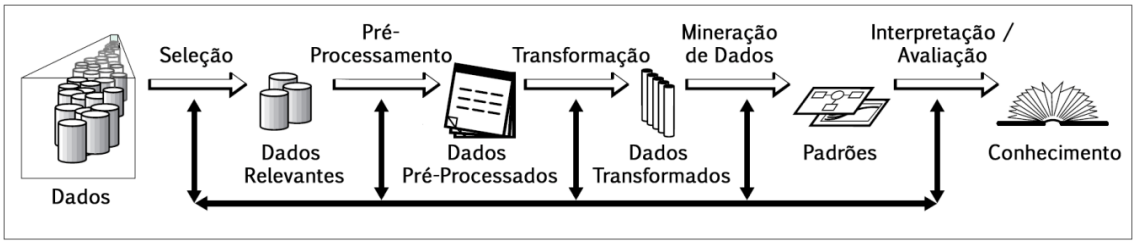


Figura 2.6 – Processo KDD.
Fonte: Modificado de Fayyad (1996).

Um dos pacotes de aplicativos mais difundidos na comunidade científica para a aplicação de técnicas de MD é o *Waikato Environment for Knowledge Analysis* - WEKA (WITTEN; FRANK, 2005), um *software* gratuito e de código aberto (*opensource*). Desenvolvido pela Universidade de Waikato (Nova Zelândia), em linguagem Java, o WEKA consiste em uma coleção de algoritmos de aprendizado de máquina para tarefas específicas de MD, dispondo de diversas ferramentas para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização de dados em duas dimensões.

Dentre os vários classificadores disponíveis no WEKA, encontram-se as árvores de decisão que são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem ou revelem classes baseadas nos valores de atributos de um conjunto de dados (DE'ATH; FABRICIUS, 2000). Este método apresenta uma estrutura e regras explícitas, possibilitando a identificação de atributos-chave (previsores) e maior facilidade para interpretar a relação entre a variável dependente e as variáveis independentes.

Para aplicar este método, é necessário definir previamente as classes a serem utilizadas no conjunto de treinamento, motivo pelo qual a árvore de decisão é conhecida como um método de classificação supervisionada. A partir de um conjunto de treinamento, a árvore de decisão é organizada em uma hierarquia de nós internos e externos que são interligados por ramos. O nó interno (não-folha) recebe como rótulo um dos atributos previsores e é considerado uma unidade de tomada de decisão, pois avalia através de teste lógico qual será o próximo nó descendente. Dependendo do resultado do teste lógico, a árvore se ramifica e este procedimento é repetido até que um nó externo seja alcançado. A repetição deste procedimento caracteriza a recursividade da árvore de decisão (BREIMAN et al., 1984). O teste lógico que determina a ramificação da árvore é baseado na entropia (medida da falta de homogeneidade dos dados de entrada em relação à sua classificação). O valor do limiar utilizado como critério para ramificação é rotulado para cada ramo que conecta um nó interno

ao seu descendente. Por fim, o nó externo (sem descendentes), também conhecido como folha, é rotulado com uma classe.

O WEKA dispõe dos seguintes algoritmos na categoria árvore de decisão: J48, BFTree, Decision Stump, FT, LADTree, LMT, NBTree, Random Forest, Random Tree, REPTree e Simple Cart. A Figura 2.7 apresenta um exemplo da classificação utilizando árvores de decisão. No gráfico da Figura 2.7a, as duas classes (vinho e verde), representadas pelos atributos x e y, são perfeitamente separadas pela combinação de partições ortogonais geradas por uma árvore de decisão, cujas regras estão especificadas na Figura 2.7b.

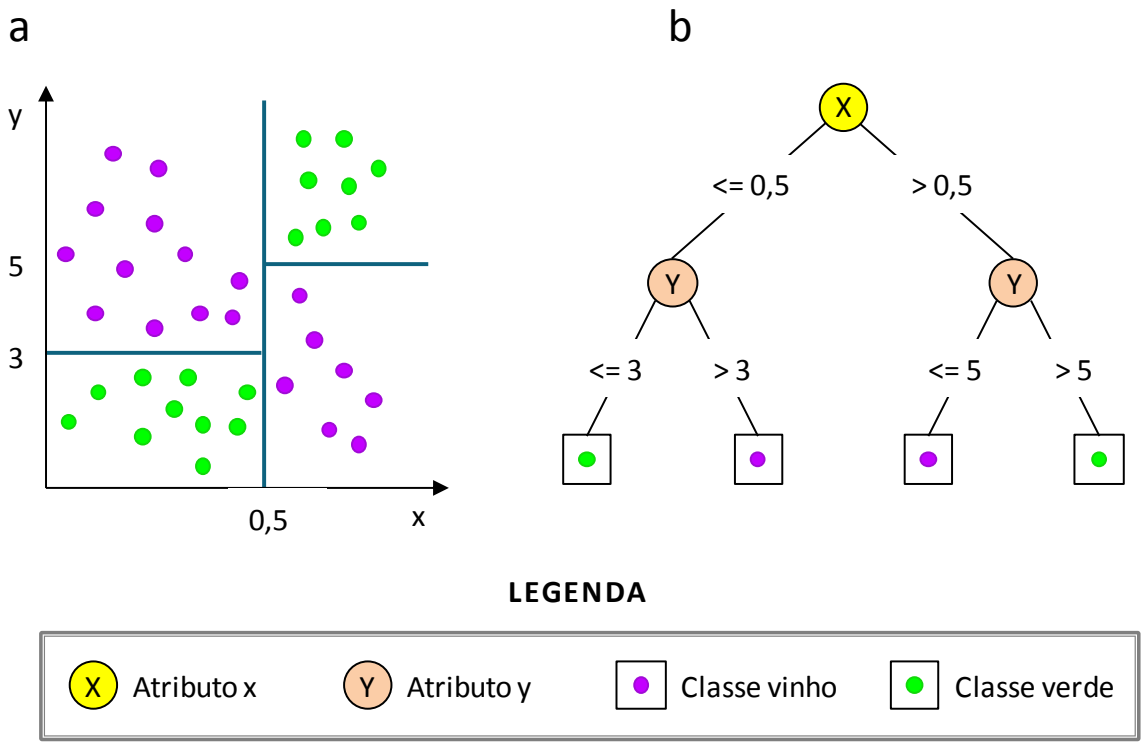


Figura 2.7 – Separação perfeita de duas classes no espaço de atributos através da combinação de partições ortogonais (a) usando a árvore de decisão (b).

3 MATERIAL E MÉTODOS

3.1. Área de estudo

A área de estudo compreende a região amazônica brasileira, na qual foram selecionadas cinco regiões com diferentes padrões geomorfológicos, cujas características constam na Tabela 3.1 e são ilustradas na Figura 3.1. A área utilizada para a aplicação da metodologia proposta (Figura 3.1a) foi denominada como área de treinamento. Esta área foi escolhida por apresentar pouca influência antrópica (como a presença de manchas urbanas, rodovias e desmatamentos), o que poderia prejudicar algumas análises devido às limitações dos dados utilizados neste trabalho. As demais regiões (Figuras 3.1b, 3.1c, 3.1d), denominadas como Áreas Teste, foram escolhidas para avaliar a aplicação do resultado da metodologia proposta.

Tabela 3.1 – Características das áreas escolhidas para os quatro estudos de caso.

Área	Retângulo Envolvente	Tamanho (pixels)	Variação Altimétrica(m)	Figura
A	0° 58' 33"S, 59° 3' 36"W a 1° 43' 33"S, 57° 48' 36"W	[1500x900]	14 a 303	11a
B	1° 0' 0"N, 69° 0' 0"W a 0° 0' 0", 68° 0' 0"W	[1200x1200]	55 a 461	11b
C	2° 4' 30"S, 59° 42' 45"W a 2° 37' 0"S, 59° 10' 15"W	[650x650]	29 a 157	11c
D	4° 1' 33"S, 60° 31' 24"W a 4° 46' 33"S, 59° 36' 24"W	[1100x900]	9 a 64	11d
E	4° 0' 0"S, 65° 0' 0"W a 5° 0' 0"S, 64° 0' 0"W	[1200x1200]	22 a 100	11e

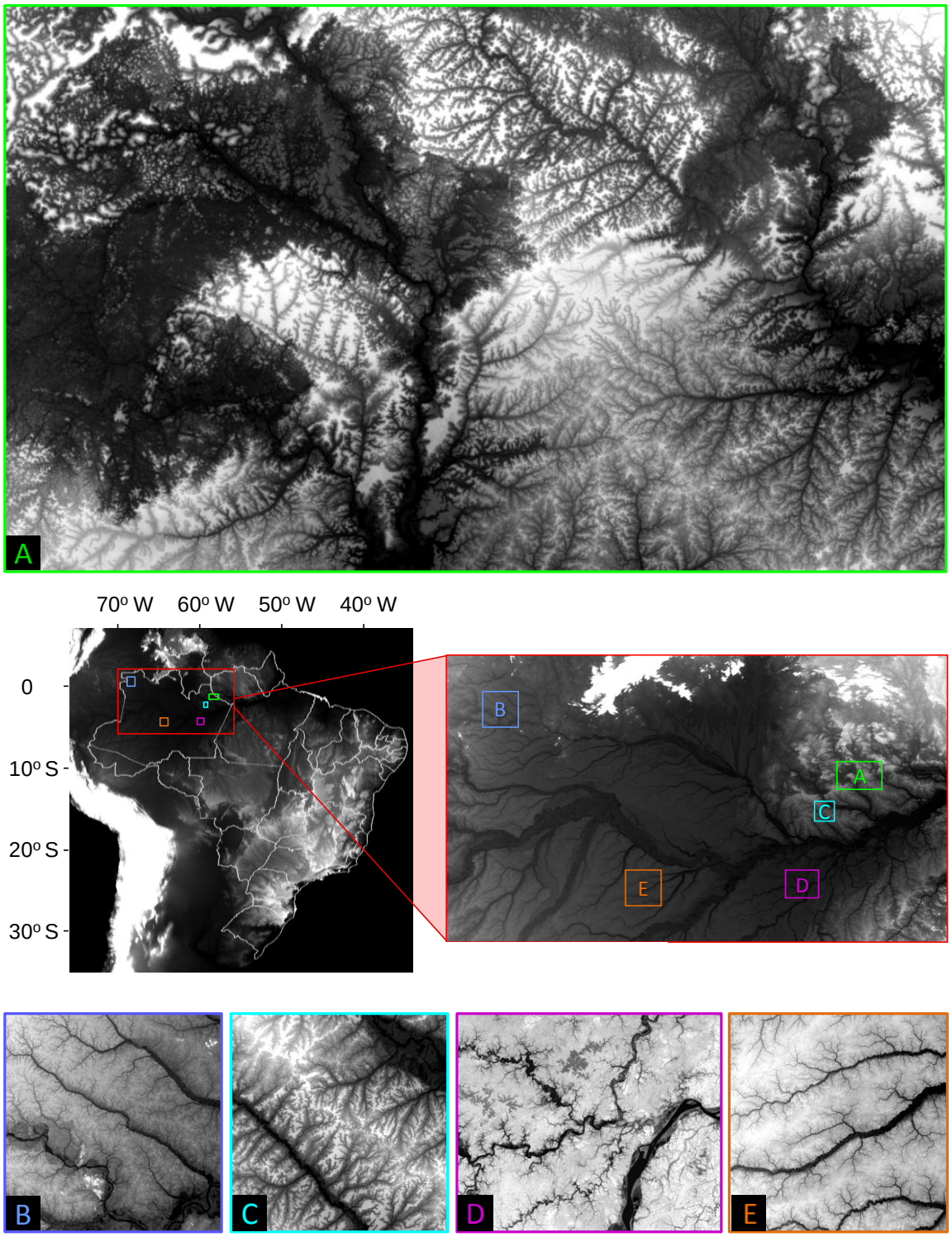


Figura 3.1 – Estudos de caso selecionados na região amazônica, sendo uma área escolhida para a aplicação da metodologia proposta (Área A) e quatro áreas para a elaboração dos testes (Áreas B, C, D e E).

3.2. Material

Para este trabalho, serão utilizados os dados altimétricos provenientes da missão SRTM. Estes dados estão disponíveis no seguinte endereço: http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/.

As quadrículas de cada região escolhida foram previamente selecionadas e os seus arquivos, originalmente no formato HGT (extensão do termo em inglês *HeiGhT,* referindo-se ao formato padrão de dados altimétricos do SRTM), foram mosaicados e convertidos para o formato Geotiff, utilizando o pacote ENVI (*The Environment for Visualizing Images*). As características originais dos dados foram mantidas: projeção geográfica, datum horizontal e vertical WGS-84, e resolução espacial de 3 arcos de segundos (aproximadamente 90 m próximo a linha do Equador).

Para a execução dos experimentos, foi desenvolvido um protótipo utilizando a linguagem de programação IDL versão 8.2.1 (*Interactive Data Language*) que integrou as funcionalidades do ENVI (versão 5.0) aos recursos do pacote de aplicativos WEKA (versão 3.6.9) em um único sistema (BANON e RENNÓ, 2013). O WEKA encontra-se disponível gratuitamente pela internet (http://www.cs.waikato.ac.nz/ml/weka/downloading.html) e foi desenvolvido utilizando a linguagem de programação Java. A tecnologia *IDL Connectivity Bridges* (EXELIS, 2010) tornou possível a manipulação de objetos desenvolvidos em Java a partir da linguagem IDL. Esta tecnologia proporciona o compartilhamento de informações entre diferentes ambientes ou linguagens de programação.

3.3. Metodologia

A metodologia geral é apresentada no fluxograma da Figura 3.2. A metodologia clássica (baseado em um único atributo) será comparada á metodologia proposta (utilizando vários atributos) para a extração automática de redes drenagens.

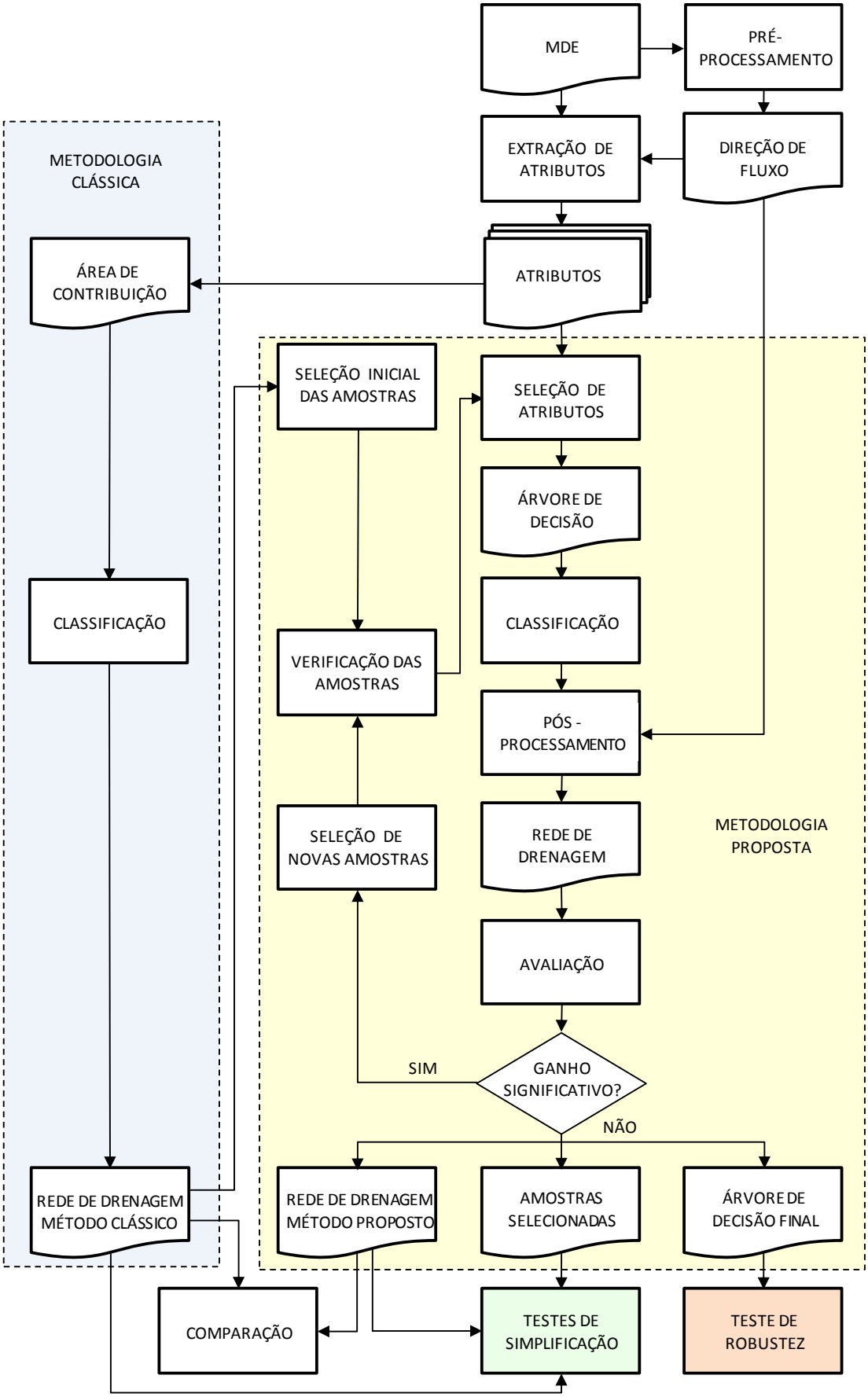


Figura 3.2 – Fluxograma geral da metodologia.

3.3.1. Pré-processamento e Extração de Atributos

O MDE SRTM foi pré-processado utilizando o Sistema TerraHidro (ROSIM et al., 2003), fornecendo dois produtos hidrologicamente consistidos: o LDD e o MDE. Para este trabalho, serão utilizados como dados de entrada, o LDD hidrologicamente consistido e o MDE original. O MDE hidrologicamente consistido será descartado, por não preservar as características morfométricas do relevo, que são fundamentais para este trabalho.

Em seguida, foram calculados os dois grupos de atributos: morfométricos e os baseados na direção de fluxo. O primeiro grupo é composto por atributos de terreno extraídos diretamente do MDE, enquanto o segundo grupo é obtido a partir da grade LDD e do MDE. Para este trabalho, foram calculados 32 atributos, sendo 18 morfométricos (Tabela 3.2) e 14 baseados na direção de fluxo (Tabela 3.3). Os atributos morfométricos são obtidos a partir da topografia local através do ajuste de uma função bivariada quadrática utilizando os valores do MDE SRTM considerando uma janela 3x3. Maiores detalhes sobre o cálculo e definições destes atributos podem ser encontrados no Apêndice A.

Tabela 3.2 – Lista de atributos morfométricos.

	Atributos	Identificador	Unidade	Fonte
Morfométricos	Declividade	slope	%	Wood (1996); Shary et al. (2002)
	Fator de Gradiente	gradfactor	$10^2 m.m^{-1}$	Shary et al. (2002)
	Curvatura Horizontal	horcurv	$10^3 m^{-1}$	Shary et al. (2002)
	Curvatura Plana	plancurv	$10^3 m^{-1}$	Wood (1996); Shary et al. (2002)
	Curvatura Vertical	vertcurv	$10^3 m^{-1}$	Wood (1996), Shary et al. (2002)
	Curvatura Média	meancurv	$10^3 m^{-1}$	Wood (1996), Shary et al. (2002)
	Não Esfericidade	unsphericity	$10^3 m^{-1}$	Shary et al. (2002)
	Diferença de curvatura	difcurv	$10^3 m^{-1}$	Shary et al. (2002)
	Rotor	rotor	$10^3 m^{-1}$	Shary et al. (2002)
	Excesso de Curvatura Horizontal	khe	$10^3 m^{-1}$	Shary et al. (2002)
	Excesso de Curvatura Vertical	kve	$10^3 m^{-1}$	Shary et al. (2002)
	Curvatura Mínima	kmin	$10^3 m^{-1}$	Wood (1996); Shary et al. (2002)
	Curvatura Máxima	kmax	$10^3 m^{-1}$	Wood (1996); Shary et al. (2002)
	Curvatura Gaussiana Total	totgausscurv	$10^6 m^{-2}$	Shary et al. (2002)
	Curvatura Circular Total	totringcurv	$10^3 m^{-2}$	Shary et al. (2002)
	Curvatura de Acumulação Total	totaccumcurv	$10^6 m^{-2}$	Shary et al. (2002)
	Curvatura Longitudinal	longcurv	$10^3 m^{-1}$	Wood (1996)
	Curvatura Transversal	crossseccurv	$10^3 m^{-1}$	Wood (1996)

Tabela 3.3 – Lista de atributos baseados na direção de fluxo.

	Atributos	Identificador	Unidade	Fonte
Baseados na Direção de Fluxo	Área de Contribuição	areacontr	m^2	Banon e Rennó (2013)
	Desnível ao Topo	desvtopo	m	Banon e Rennó (2013)
	Desnível Vertical	desvvert	m	Banon e Rennó (2013)
	Declividade à Jusante	decld8	%	Banon e Rennó (2013)
	Ordem Máxima de Sthraler	ordstrmax	adimensional	Banon e Rennó (2013)
	HAND com Ordem Máxima de Sthraler 3	hand_ordstrmax3	m	Banon e Rennó (2013)
	HAND com Ordem Máxima de Sthraler 4	hand_ordstrmax4	m	Banon e Rennó (2013)
	HAND com Ordem Máxima de Sthraler 5	hand_ordstrmax5	m	Banon e Rennó (2013)
	HAND com Ordem Máxima de Sthraler 6	hand_ordstrmax6	m	Banon e Rennó (2013)
	HAND com Ordem Máxima de Sthraler 7	hand_ordstrmax7	m	Banon e Rennó (2013)
	Desnível ao 1º vizinho à jusante	difz1	m	Banon e Rennó (2013)
	Desnível ao 4º vizinho à jusante	difz4	m	Banon e Rennó (2013)
	Desnível ao 16º vizinho à jusante	difz16	m	Banon e Rennó (2013)
	Desnível ao 64º vizinho à jusante	difz64	m	Banon e Rennó (2013)

3.3.2. Metodologia clássica

A metodologia clássica baseia-se na escolha empírica de um valor que determina a área de contribuição mínima necessária para a definição da rede de drenagem. O processo de classificação consiste na identificação e rotulação das células da grade Área de Contribuição (Tabela 3.3) cujos valores são iguais ou superiores ao limiar escolhido.

Neste trabalho, adotou-se o limiar igual a 435.000m^2, que equivale ao limiar de 50 pixels adotado por Rennó et al. (2008) em dados do MDE SRTM. Estes autores observaram que o uso deste limiar na região amazônica gerou redes de drenagem coerentes com a realidade.

3.3.3. Metodologia proposta

A metodologia proposta inicia-se com a escolha de um conjunto amostral de pontos representativos das feições características da rede de drenagem. A partir desta amostra, é feita a seleção dos atributos e a definição de critérios utilizados pelo classificador para a obtenção da rede de drenagem. Neste

trabalho, foi utilizado o algoritmo J48, um classificador do tipo árvore de decisão também conhecido como algoritmo C4.5 (QUINLAN, 1993).

3.3.3.1. Seleção inicial das amostras e verificação

Para realizar a etapa de seleção dos atributos, é necessário obter uma amostra para treinamento. A escolha dos pontos amostrais pode ser elaborada por um especialista, como apresentado por Banon et al. (2013). No entanto, nem todo experimento poderá contar com a presença de um especialista, que precisa estar dedicado em todas as etapas de seleção das amostras, uma vez que a Mineração de Dados é um processo iterativo.

Na fase de seleção das amostras, os pontos devem estar distribuídos uniformemente na grade, caracterizando as três classes em estudo: nascente, drenagem e não drenagem, como definido por Banon et al. (2013). A Figura 3.3 exemplifica os pontos típicos de cada uma das classes.

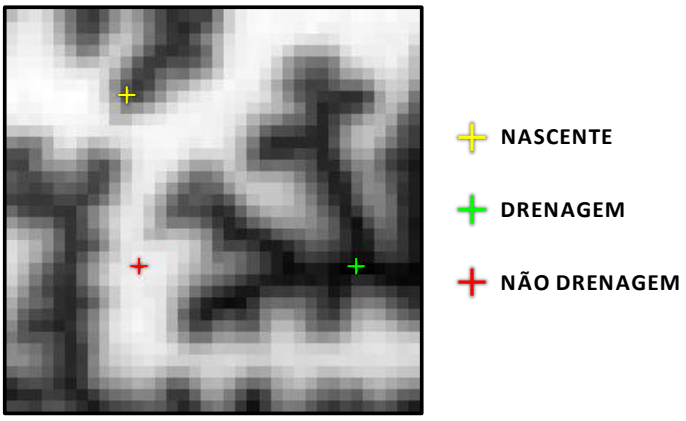


Figura 3.3 – Exemplos de pontos amostrais representativos das três classes em estudo sobre o MDE SRTM.

Inicialmente, alguns testes foram realizados com a coleta manual das amostras sem a utilização de qualquer recurso, mas este procedimento se mostrou inviável para a obtenção de uma amostra de treinamento com uma distribuição uniforme. Para minimizar este problema, os pontos amostrais foram sorteados

aleatoriamente sobre uma rede de drenagem prévia. Esta rede foi gerada a partir da metodologia clássica apresentada no item 3.3.2.

O número total de pontos sorteados inicialmente sobre a referida rede de drenagem não deve ser muito elevado. Isso pode ocasionar um gasto considerável de tempo para a revisão de todos os pontos, sem a garantia de formar um conjunto amostral representativo dos diversos padrões de feições, o que neste caso, pode resultar em uma rede de drenagem irreal. Em contrapartida, um tamanho reduzido para a amostra inicial permite uma cuidadosa revisão dos pontos selecionados, de modo que estes pontos sejam os mais característicos de cada classe. Nesta fase, as regiões cujas feições não são muito bem definidas devem ser evitadas.

Para este trabalho, utilizou-se a técnica dos sorteios aleatórios de uma pequena amostra constituída por 40 pontos, sendo 10 da classe de nascente, 10 da classe de drenagem e 20 da classe de não drenagem. Considerando que a rede de drenagem obtida pela metodologia clássica pode apresentar inconsistências, é necessário que todos os pontos amostrados sejam revisados e, nos casos onde a metodologia clássica apresentar falhas, as amostras devem ser reposicionadas para um ponto cuja feição seja característica da classe em questão.

3.3.3.2. Seleção de atributos, classificação e pós-processamento

Para a seleção de atributos foi utilizado o algoritmo J48 disponível no pacote de aplicativos WEKA. Este algoritmo gera como resultado uma árvore de decisão que, por possuir regras explícitas, possibilita uma fácil interpretação do significado de seus limiares. Além disso, a árvore de decisão permite avaliar a importância dos atributos selecionados para a etapa de classificação.

O algoritmo J48 possui uma lista de parâmetros que devem ser ajustados. A Tabela 3.4 lista os dados referentes aos parâmetros deste classificador, apresentando respectivamente: os nomes dos parâmetros, a letra utilizada para

a identificação dos parâmetros no WEKA, os significados e os valores padrões iniciais.

Tabela 3.4 – Lista de parâmetros do algoritmo J48.

Algoritmo J48			
Parâmetros	ID	Significado	Padrão
binarySplits	B	Recurso para utilizar a árvore binária.	Falso
confidenceFactor	C	Define o fator que será usado para a poda.	0,25
minNumObj	M	Define o número mínimo de amostras por folha	2
numFolds	N	Define o número de partições usadas para poda com redução de erros. Apenas uma partição é usada para poda.	3
reducedErrorPrunning	R	Recurso para realizar a poda com redução de erros	Falso
subtreeRaising	S	Recurso para utilizar a subárvore durante a poda.	Verdadeiro
unpruned	U	Recurso para desabilitar o recurso de poda	Falso
useLaplace	L	Recurso para suavização do número de folhas baseando-se em Laplace.	Falso

Em geral, os pontos amostrais são escolhidos sobre a grade MDE SRTM utilizando um Sistema de Informação Geográfica (SIG) e os valores dos atributos são exportados para um arquivo ARFF (sigla do inglês *Attribute-Relation File Format*). Em seguida, os dados contidos neste arquivo são importados para o pacote de aplicativos WEKA, com o qual é realizada a fase de treinamento, resultando em uma árvore de decisão. Através do desenvolvimento de um aplicativo, as regras contidas na árvore de decisão podem ser aplicadas em cada célula das grades de atributos, gerando a rede de drenagem. A linguagem Java seria a mais indicada para criar este aplicativo, uma vez que o WEKA também foi desenvolvido em Java, mas outras linguagens ou até mesmo ferramentas de SIG poderiam ser utilizadas para este propósito (Figura 3.4a).

Para facilitar a reprodutibilidade dos experimentos realizados neste trabalho, evitando a manipulação de arquivos externos, um protótipo denominado *DrainageDataMining* (BANON; RENNÓ, 2013) foi desenvolvido com uma

interface gráfica associada às funcionalidades do pacote de aplicativos WEKA. Neste contexto, utilizando a linguagem IDL, foram integradas as etapas de coleta de amostras, armazenamento de valores de cada ponto e seleção de atributos. A fase de classificação propriamente dita, também foi incorporada ao protótipo (Figura 3.4b).

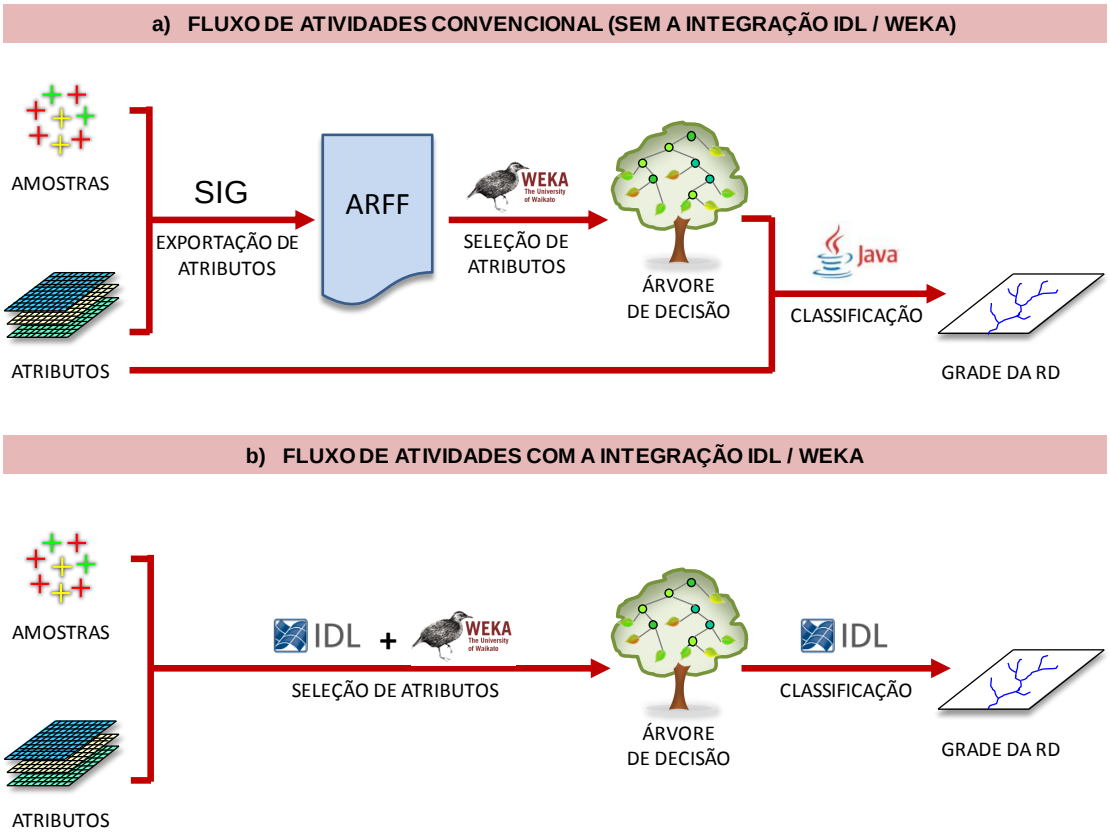


Figura 3.4 – Fluxo de atividades: a) Convencional e b) Proposto.

A partir do WEKA, foram obtidas as regras da árvore de decisão, mas a execução destas regras no protótipo foi desenvolvida de forma independente. O intuito deste procedimento foi a otimização da etapa de classificação. A partir da árvore obtida pelo WEKA, cada caminho partindo do nó principal em direção a uma folha, tornou-se uma regra booleana (Figura 3.5). A aplicação destas regras booleanas sobre a grade de atributos resultou na rede de drenagem.

A etapa de classificação não garante a continuidade da rede de drenagem, demandando uma etapa de pós-processamento. Como o LDD indica a direção

de fluxo, é possível encontrar os caminhos que conectam os trechos isolados à rede de drenagem principal (Figura 3.6).

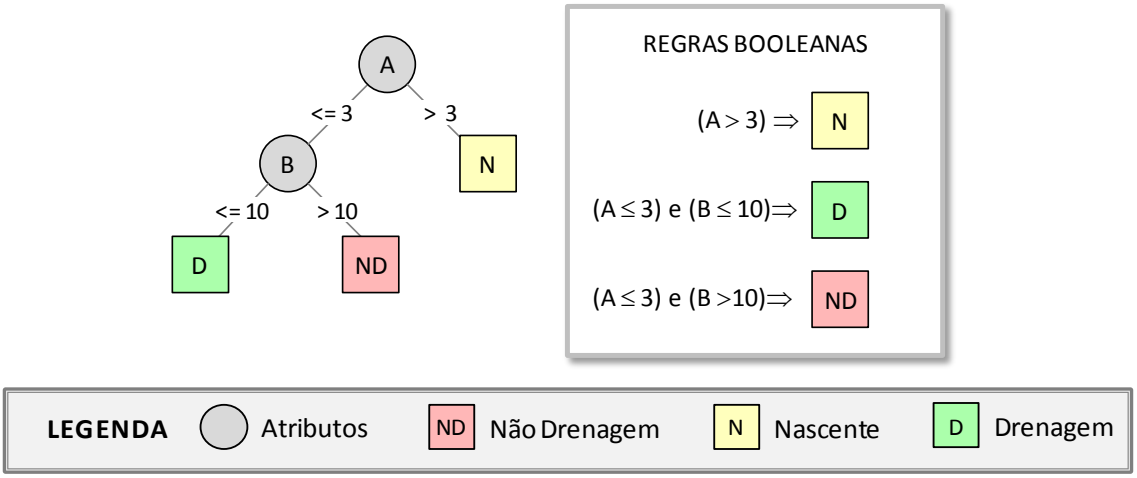


Figura 3.5 – Definição de regras booleanas a partir de uma árvore de decisão.

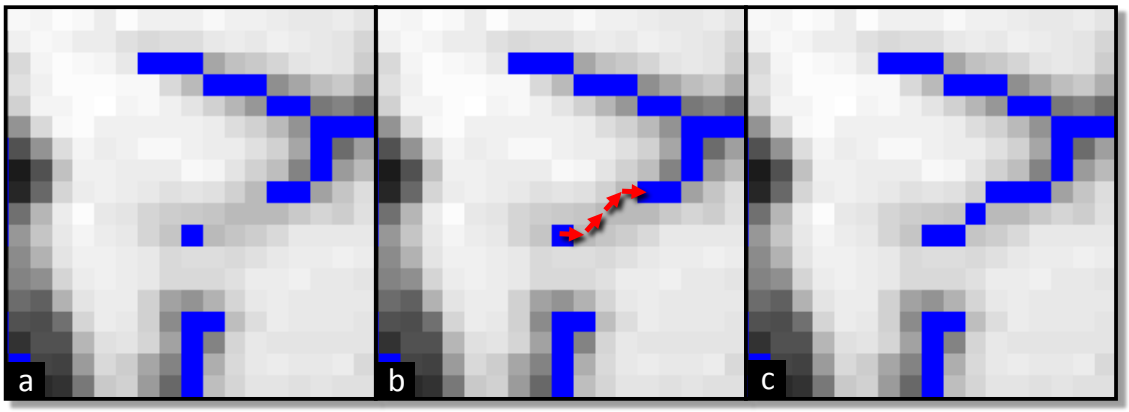


Figura 3.6 – Rede de drenagem (em azul) sobre o MDE SRTM: a) antes do pós-processamento, b) indicação da direção de fluxo (em vermelho) e c) resultado do pós-processamento.

Quando os trechos isolados caracterizam uma feição hidrológica, a sua conexão à rede de drenagem é necessária. No entanto, quando o MDE SRTM apresenta regiões antropizadas ou com diferentes usos e coberturas, é comum observar algumas manchas que resultam na classificação de pontos isolados ou pequenos trechos distantes da rede de drenagem. Estes pontos não caracterizam uma feição hidrológica e poderiam resultar em um trabalho considerável de edição. Neste trabalho, é proposta uma etapa de pós-

processamento que permite a definição de dois parâmetros para o ajuste fino da rede de drenagem. A seguir são apresentados os três passos para este ajuste, que estão exemplificados na Figura 3.7:

a) os trechos isolados com intervalo (número de células da grade) menor ou igual a um determinado valor, serão conectados (definição do primeiro parâmetro chamado *gapsize*);
b) os trechos isolados com tamanho inferior ou igual a um valor pré-definido serão apagados (definição do segundo parâmetro chamado *segsize*); e
c) todos os trechos isolados são conectados.

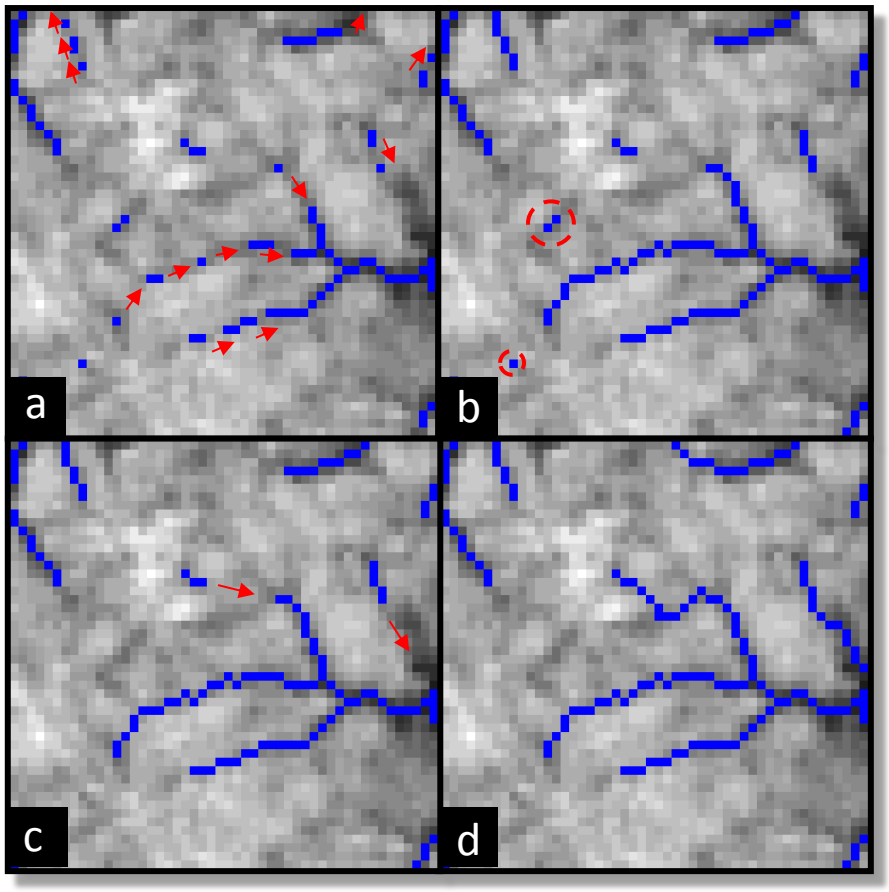


Figura 3.7 – Rede de drenagem obtida pela metodologia proposta: a) antes do pós-processamento; b) após a correção dos intervalos com até 5 células; c) após a eliminação de trechos isolados com menos de 3 células (destacados em vermelho em b) e d) conexão de trechos isolados.

3.3.3.3. Avaliação da rede de drenagem e seleção de novas amostras

Inicialmente, a avaliação consistiu na análise da matriz de confusão resultante do processo de construção da árvore de decisão utilizando o algoritmo J48 (produzida pelo WEKA). Uma atenção especial deve ser dedicada aos erros entre as classes não drenagem e nascente; e não drenagem e drenagem. Como as classes de nascente e drenagem são partes da rede de drenagem, a confusão entre estas duas classes deve ser desconsiderada.

Durante a fase de revisão das amostras, a partir do protótipo é possível listar os pontos classificados incorretamente e visualizá-los no MDE SRTM. Ao detectar amostras não representativas, a sua localização deve ser alterada para posições mais precisas, sobre feições bem características. Nos casos de amostras bem representativas, mas classificadas incorretamente, mantém-se a posição da amostra e selecionam-se novos pontos com as mesmas características. O intuito deste procedimento é permitir que o classificador aprenda a reconhecer esta nova feição ainda incomum na amostra.

Nesta etapa de revisão dos pontos, a porcentagem de acerto da classificação fornecida pelo WEKA não foi avaliada, pois as altas porcentagens de acerto nem sempre resultam em uma rede de drenagem adequada. Isso ocorre porque se o número de erros for reduzido, mas os poucos erros forem pontos de nascente ou drenagem distantes de um grande rio, o pós-processamento conectará estes pontos resultando em uma rede de drenagem irreal. Por outro lado, se o número de erros for alto, mas a confusão estiver entre as classes de nascente e drenagem, estes erros não causarão impacto na rede de drenagem pós-processada.

A etapa de avaliação também considerou a estrutura da árvore de decisão. Neste caso, foram observados os atributos selecionados pela árvore, o valor dos limiares, a interpretação do significado hidrológico desta estrutura e, em especial, a análise de suas folhas. As folhas mais inexpressivas da árvore, ou seja, aquelas com um número reduzido de amostras ou com muitos erros

indicam a necessidade da análise de sua representatividade em relação à área de estudo. Em geral, estas folhas inexpressivas estão associadas a feições muito particulares e podem resultar em vários pontos inconsistentes gerando uma rede de drenagem irreal.

Para avaliar a representatividade de cada folha da árvore em relação à área de estudo, foi desenvolvido o recurso de classificação por folha, originalmente inexistente no pacote de aplicativos do WEKA. Aplicando-se as regras booleanas nas grades dos atributos, obteve-se uma rede de drenagem, onde cada célula da grade possui a identificação da folha da árvore na qual ocorreu a sua classificação. A partir deste conceito, recursos gráficos do protótipo puderam indicar as folhas responsáveis pelos erros de classificação, como os falsos trechos de drenagem ou a ausência de feições hidrológicas. Este recurso também foi fundamental para auxiliar na escolha de novas amostras.

Uma vez constatada a presença de uma folha inexpressiva contendo erros de classificação, o protótipo também permitiu identificar qual era a amostra responsável por cada erro desta folha. Em seguida, esta amostra era visualizada sobre o MDE para analisar a sua coerência em relação à feição observada. A partir do resultado da interpretação, a amostra poderia ser reposicionada, excluída ou até mesmo preservada, mas neste último caso o padrão da feição precisava ser reafirmado, selecionando novos pontos com as mesmas características e pertencentes à mesma folha da amostra incorreta.

O processo de coleta de novas amostras pode ocorrer simultaneamente à etapa de avaliação dos resultados da classificação ou através de um sorteio aleatório de pontos sobre a rede de drenagem obtida. No caso do sorteio, todos os pontos devem ser revisados a fim de garantir a sua coerência. Caso os pontos sejam coletados durante a avaliação, pode-se utilizar a rede de drenagem apenas como um referencial para identificar falsos trechos de drenagem ou a ausência de feições hidrológicas. A seleção de novas amostras implica que a etapa de seleção de atributos seja refeita, resultando em uma nova árvore de decisão, que por sua vez gera uma nova rede de drenagem.

Estas etapas são repetidas até que a nova rede de drenagem não apresente um ganho significativo em relação à última classificação. Desta forma, a proporção do número de amostras escolhida inicialmente para cada classe pode sofrer alteração.

A avaliação deste ganho significativo pode ser feito através de uma avaliação qualitativa, na qual os critérios são baseados na interpretação visual, buscando o equilíbrio entre a presença de falsas feições e a ausência de feições que deveriam estar presentes na rede de drenagem obtida. Outra forma de avaliação consiste na análise do ganho significativo, através de um teste estatístico baseado na comparação de áreas aleatoriamente escolhidas sobre o MDE. Definiu-se que estas áreas teriam pelo menos 80x80 células, para garantir que a informação contextual estivesse representada. Em cada área, foi sobreposta cada uma das duas redes de drenagem a serem comparadas (anterior e atual), escolhendo-se então aquela cuja rede de drenagem é mais representativa de acordo com uma avaliação qualitativa. Neste trabalho, definiu-se como 50 o número total de áreas avaliadas por experimento (n). Este número foi escolhido levando-se em consideração o tempo gasto para executar uma avaliação e o número total de experimentos realizados.

A hipótese nula do teste de ganho significativo considera que, caso as duas classificações resultem em redes de drenagem semelhantes, espera-se a casualidade na escolha (probabilidade $p = 0,5$), ou seja, que não haja preferência de uma determinada rede de drenagem. Supondo-se que X represente o total de áreas da rede de drenagem atual considerada como melhor, então X terá uma distribuição binomial com parâmetros $n = 50$ e $p = 0,5$ se a hipótese nula for verdadeira. Se x representar o número efetivamente observado numa determinada avaliação, então $P = P(X \geq x)$ representa a probabilidade de se escolherem x ou mais áreas de forma totalmente casual. Dessa forma, se $P <= 5\%$ pode-se rejeitar a hipótese nula, concluindo-se que houve melhora significativa em relação à rede de drenagem anterior. Este teste é similar ao “Teste dos Sinais” (SIEGEL, 1977). Para a realização deste teste,

utilizou-se a avaliação cega, desenvolvendo-se uma ferramenta para o sorteio e a seleção do melhor resultado pelo avaliador, de modo que o mesmo não tivesse o conhecimento prévio da origem das redes de drenagem em questão. Para permitir a reprodutibilidade dos experimentos realizados neste trabalho, os dados utilizados em cada teste comparativo e seu resultado foram disponibilizados em Banon (2013).

3.3.4. Comparação entre as redes de drenagens obtidas pelas metodologias clássica e proposta

Inicialmente, a comparação dos resultados das duas metodologias pode seguir a mesma abordagem adotada no processo de avaliação de uma rede de drenagem a partir da análise do ganho significativo apresentada no item **3.3.3.3**. Neste caso, uma rede de drenagem representa a rede obtida pela metodologia clássica e a outra pela metodologia proposta. Espera-se que a nova metodologia seja significativamente superior à clássica, uma vez que a metodologia proposta utiliza vários atributos incluindo a área de contribuição.

Adicionalmente, esta comparação também pode incluir uma análise a partir do cruzamento das duas redes de drenagem. Neste contexto, são identificados os segmentos comuns às duas redes e aqueles que aparecem exclusivamente em cada uma delas. Através de uma avaliação qualitativa, é possível identificar se uma feição representada por um determinado segmento é real ou não. É importante observar que as feições não detectadas por ambas as metodologias, assim como, as feições erroneamente detectadas pelas duas metodologias não são consideradas nesta abordagem.

3.3.5. Testes de Simplificação

Neste trabalho, propõe-se testar três procedimentos com o intuito de verificar a possibilidade de simplificação da metodologia proposta, uma diminuindo o número de classes que representam as feições hidrológicas da rede de drenagem, e duas restringindo o número de atributos a serem considerados durante a etapa de seleção dos mesmos.

Nos três procedimentos, após a seleção dos atributos, a árvore de decisão é gerada e utilizada na etapa de classificação da grade de atributos obtendo-se assim, a rede de drenagem. Cada uma das redes é então comparada à rede de drenagem resultante da metodologia proposta, que utiliza todos os atributos e as três classes definidas inicialmente. A comparação das redes de drenagem deve seguir a mesma abordagem do item 3.3.4. No fluxograma da Figura 3.8, são apresentadas as etapas realizadas nos três testes de simplificação.

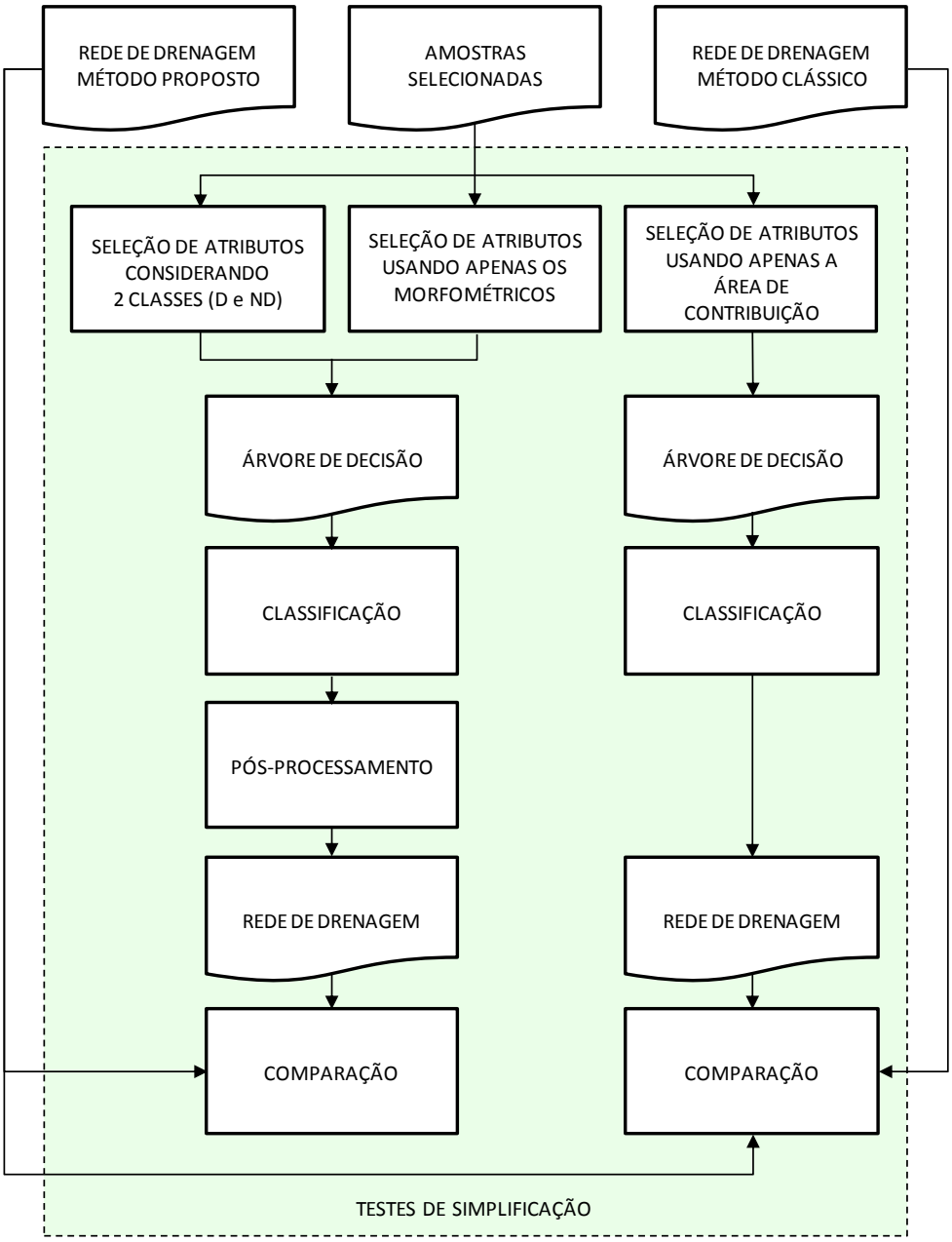


Figura 3.8 – Fluxograma do Teste de Simplificação.

3.3.5.1. Simplificação do número de classes

A metodologia proposta baseia-se na seleção de amostras de 3 classes previamente definidas: nascente, drenagem e não drenagem. As feições relacionadas às nascentes pertencem à rede de drenagem e sendo assim, os pontos de nascente e drenagem poderiam ser agrupadas em uma única classe. Dessa forma, os erros de confusão entre estas duas classes seriam desconsiderados. Por outro lado, a inclusão da classe nascente pode ser importante para definir o tênue limite do início da rede de drenagem durante a etapa de seleção de amostras, permitindo que o classificador reconheça o padrão de suas potenciais ocorrências.

Neste contexto, considera-se a hipótese de que a árvore de decisão possa ser simplificada, sem comprometer o resultado da classificação. Para realizar este teste, a partir do conjunto amostral obtido pela metodologia proposta, os pontos da classe nascente foram incorporados à classe drenagem. Desta forma, o conjunto amostral apresenta apenas duas classes: drenagem e não drenagem.

Na avaliação dos resultados, devem ser observados os seguintes aspectos: a complexidade e o significado hidrológico da árvore de decisão, os erros de classificação, a representatividade das folhas e a qualidade da rede de drenagem resultante.

3.3.5.2. Simplificação com a utilização exclusiva dos atributos morfométricos

A metodologia proposta utiliza 18 atributos morfométricos e 14 baseados na direção de fluxo. Para calcular os atributos morfométricos, não é necessário realizar a etapa de pré-processamento do MDE. Sendo assim, a simplificação da metodologia proposta utilizando apenas atributos morfométricos, resultaria em uma economia de tempo, dispensaria o uso de sistemas especializados para o pré-processamento e evitaria os problemas que podem ocorrer durante esta etapa, como por exemplo, a criação de artefatos resultantes da eliminação dos sumidouros.

Neste contexto, considera-se a hipótese de que estes atributos sejam suficientes para representar as feições hidrológicas associadas à rede de drenagem. Espera-se então que a rede de drenagem obtida seja semelhante àquela na qual todos os atributos são utilizados (metodologia proposta).

3.3.5.3. Simplificação através da definição do limiar ótimo para a área de contribuição

A área de contribuição é o atributo comumente utilizado na extração automática de redes de drenagem. Usualmente o valor do limiar é definido arbitrariamente, baseando-se na literatura ou através de um teste empírico fundamentado em critérios de interpretação visual.

A partir de um conjunto amostral representativo das feições hidrológicas, formulou-se a hipótese de que é possível definir um limiar ótimo utilizando a mineração de dados, ou seja, um valor que minimize os erros de classificação. Espera-se que o resultado da aplicação deste limiar supere o obtido pela metodologia clássica, que se baseia no valor sugerido por Rennó et al. (2008).

Para a definição de um único limiar a partir da aplicação de um classificador do tipo árvore de decisão, é necessário utilizar apenas 2 classes. Caso fossem utilizadas as 3 classes propostas neste trabalho, a árvore teria pelo menos três folhas, o que implicaria na obtenção de dois limiares para esta diferenciação. Desta forma, a fim de comparar o limiar adotado pela metodologia clássica com o valor ótimo obtido a partir da Mineração de Dados, nesta abordagem os pontos da classe nascente foram agrupados aos da classe drenagem, definindo-se assim apenas duas classes: drenagem e não drenagem.

Caso a comparação do resultado da mineração de dados usando apenas a área de contribuição seja superior ao da metodologia clássica, a sua comparação com o resultado da metodologia proposta também será realizada.

3.3.6. Testes de Robustez

Inicialmente a metodologia proposta foi aplicada na área de treinamento, apresentada na Figura 3.1a. Esta região foi escolhida devido a sua grande diversidade de padrões geomorfológicos, o que sugere a obtenção de um conjunto amostral bem representativo. Supõe-se, portanto, que este conjunto amostral poderia gerar uma árvore de decisão robusta, ou seja, extrapolável para outras regiões, e ainda assim, obter resultados superiores à metodologia clássica.

Para testar esta hipótese, foram selecionadas quatro áreas da região amazônica (Figuras 3.1b, 3.1c, 3.1d e 3.1e). A grade de atributos de cada área foi gerada a partir do MDE SRTM, seguindo as etapas apresentadas no fluxograma da metodologia geral (Figura 3.2). As regras da árvore de decisão obtidas para a área de treinamento foram aplicadas sobre estas grades, gerando a rede de drenagem de cada área teste. Estas redes foram então comparadas às redes obtidas a partir da aplicação da metodologia clássica. A comparação das redes de drenagem seguiu a mesma abordagem descrita no item 3.3.4. O fluxograma das etapas do Teste de Robustez é apresentado na Figura 3.9.

A Mineração de Dados é um processo iterativo, que demanda a análise cuidadosa em cada etapa do processo e o tempo dispensado nas suas várias fases depende do conhecimento prévio de quem executa esta técnica. A obtenção de uma árvore robusta teria como benefício imediato, um resultado superior ao da metodologia clássica, sem a necessidade de realizar todas as etapas de mineração de dados, o que representaria um grande avanço por ser uma técnica muito mais rápida de ser utilizada.

Ao aplicar uma árvore de decisão robusta, o resultado poderia ser aprimorado com uma breve análise da representativa das folhas desta árvore na área teste, seguido de alguns ajustes na etapa de pós-processamento. O intuito desta abordagem seria a identificação de folhas genéricas e específicas. As folhas

referentes às feições específicas da área de treinamento podem ocasionar algumas classificações incorretas na área teste. Neste caso, ao visualizar o MDE SRTM da área teste com a sobreposição da classificação da folha atípica, podem ser observados alguns pontos isolados e o aparecimento de curtos trechos desconexos que não caracterizam nenhuma feição. Estas imperfeições poderiam facilmente ser corrigidas, ajustando os parâmetros da etapa de pós-processamento.

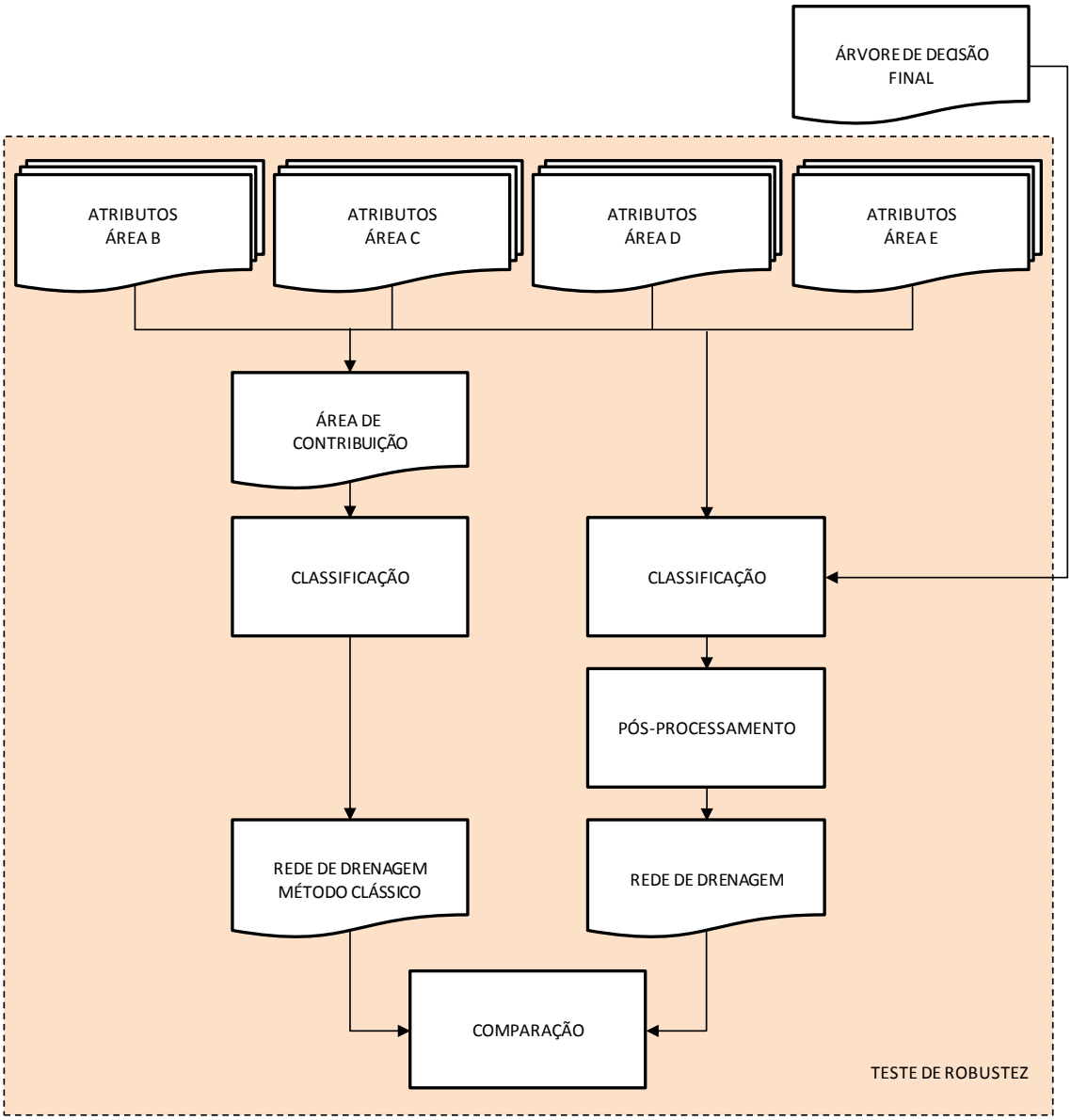


Figura 3.9 – Fluxograma do Teste de Robustez.

4 RESULTADOS E DISCUSSÃO

4.1. Rede de drenagem obtida a partir da metodologia clássica

Na metodologia clássica, a extração automática de redes de drenagem se baseia na escolha de apenas um limiar para um único atributo: a área de contribuição. Desta forma, em regiões com diferentes padrões geomorfológicos, a aplicação da metodologia clássica não deve apresentar resultados satisfatórios. A Figura 4.1 apresenta uma região heterogênea da área de treinamento (Área A), na qual foi testada a metodologia clássica adotando o limiar de 435.000 m^2. Nesta figura, é possível comparar a rede de drenagem obtida a partir da metodologia clássica, com o MDE SRTM sem a sobreposição da rede de drenagem. A área circular superior apresenta um detalhe das feições com padrões dendríticos, onde é possível observar que esta metodologia não classificou algumas pequenas drenagens. Nas duas áreas circulares inferiores, são destacadas áreas próximas a um platô, região onde esta metodologia criou falsos trechos de drenagem.

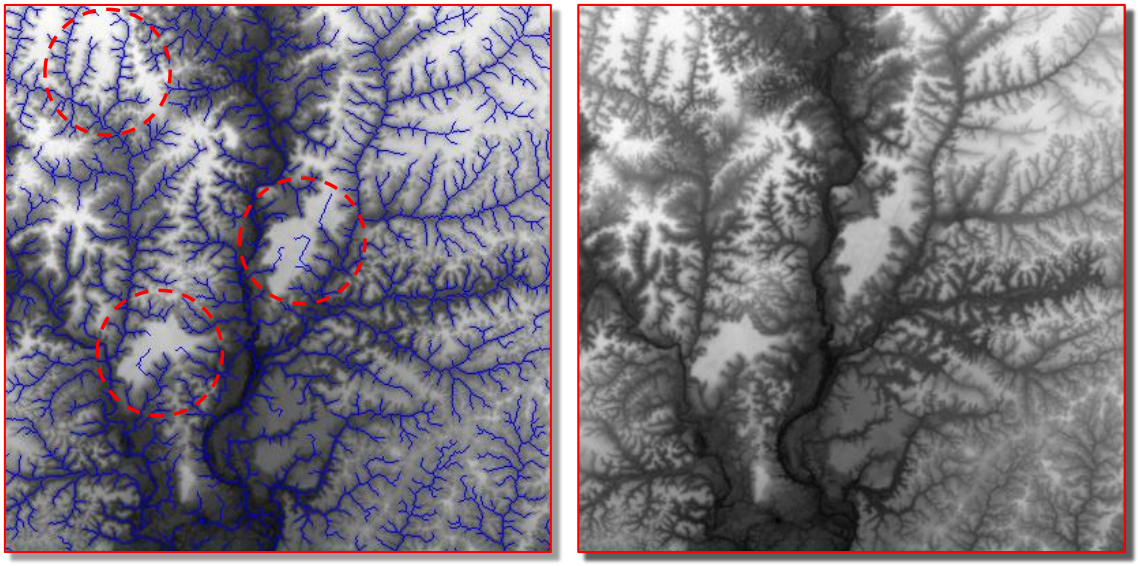


Figura 4.1 – Região heterogênea com rede de drenagem (em azul) resultante da metodologia clássica sobre o MDE SRTM.

4.2. Rede de drenagem obtida a partir da metodologia proposta

4.2.1. Fase Inicial: Seleção das Amostras, Árvore de Decisão e Avaliação

A aplicação da metodologia proposta se iniciou com a seleção de um conjunto amostral de 40 pontos (sendo 20 da classe de não drenagem, 10 de nascente e 10 de drenagem). Estes pontos foram sorteados aleatoriamente sobre a rede de drenagem gerada a partir da metodologia clássica. Como a metodologia clássica não apresenta resultados satisfatórios em áreas heterogêneas, todo o conjunto amostral foi revisado. Nesta fase, observou-se que, alguns pontos foram sorteados exatamente em posições onde a metodologia clássica falhou e, neste caso, o seu posicionamento foi corrigido. O conjunto amostral obtido após a fase de revisão é apresentado na Figura 4.2.

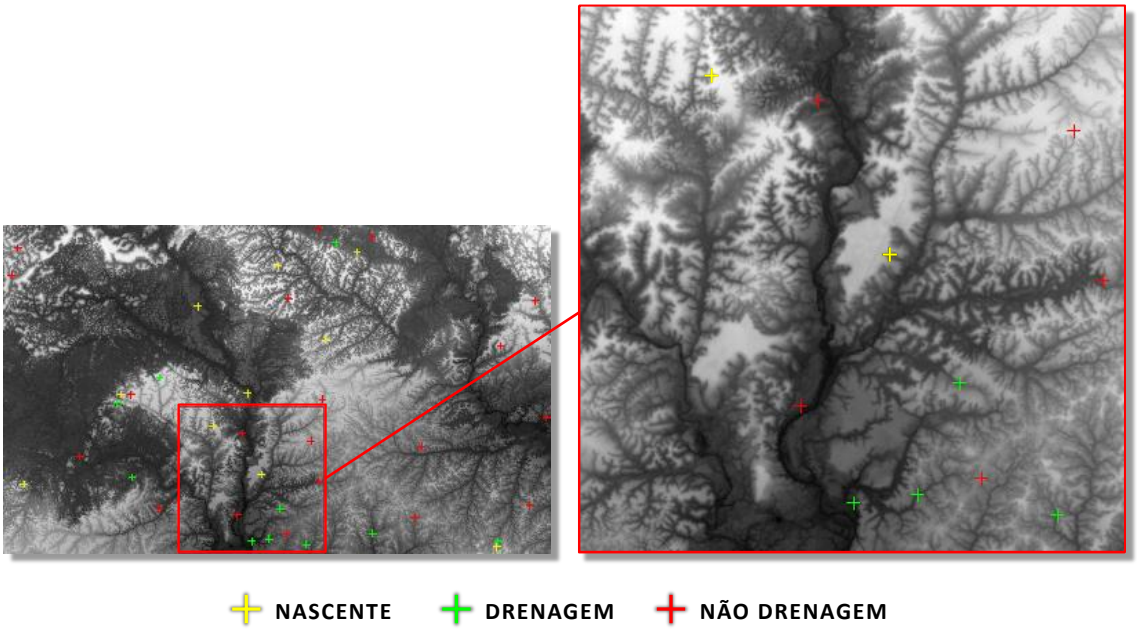


Figura 4.2 – Conjunto amostral resultante da primeira etapa de seleção das amostras após a fase de revisão dos pontos.

Em seguida, utilizando o classificador J48 foi realizada a etapa de seleção dos atributos, gerando a árvore de decisão apresentada na Figura 4.3. Nesta árvore, somente dois atributos foram selecionados: a Ordem Máxima de

Sthraler e a Área de Contribuição. Considerando que nesta primeira etapa, o reduzido conjunto amostral foi sorteado sobre uma rede de drenagem gerada a partir da Área de Contribuição, era esperado que este atributo fosse selecionado. Como há uma forte relação entre a Área de Contribuição e a Ordem Máxima de Sthraler (Figura 4.4), a sua seleção também é justificável.

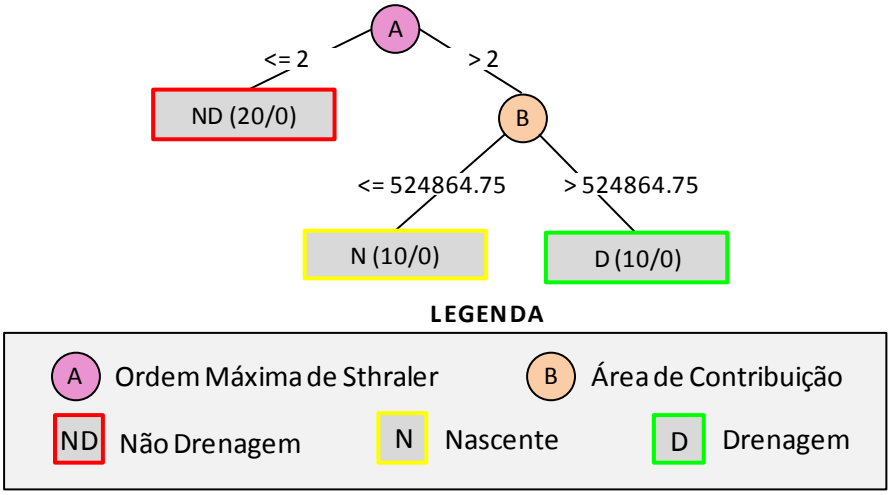


Figura 4.3 – Árvore de decisão gerada a partir do primeiro conjunto amostral. Para cada folha, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação.

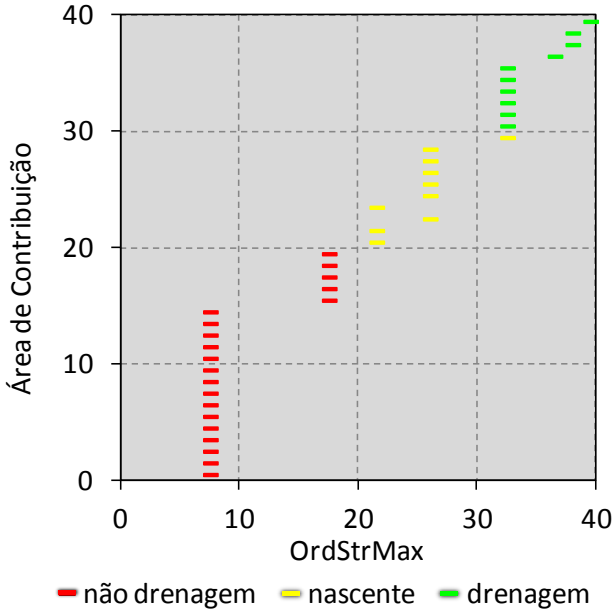


Figura 4.4 – Relação entre da Área de Contribuição e da Ordem Máxima de Sthraler (em valores ranqueados)

A partir da etapa de Pós-processamento, obteve-se a rede de drenagem apresentada na Figura 4.5. Nas regiões em destaque, é possível observar que as duas áreas circulares inferiores próximas ao platô apresentaram densos trechos de falsas drenagens. Em contrapartida, no círculo mais acima a área com padrão dendrítico foi bem classificada, incluindo até mesmo as pequenas drenagens.

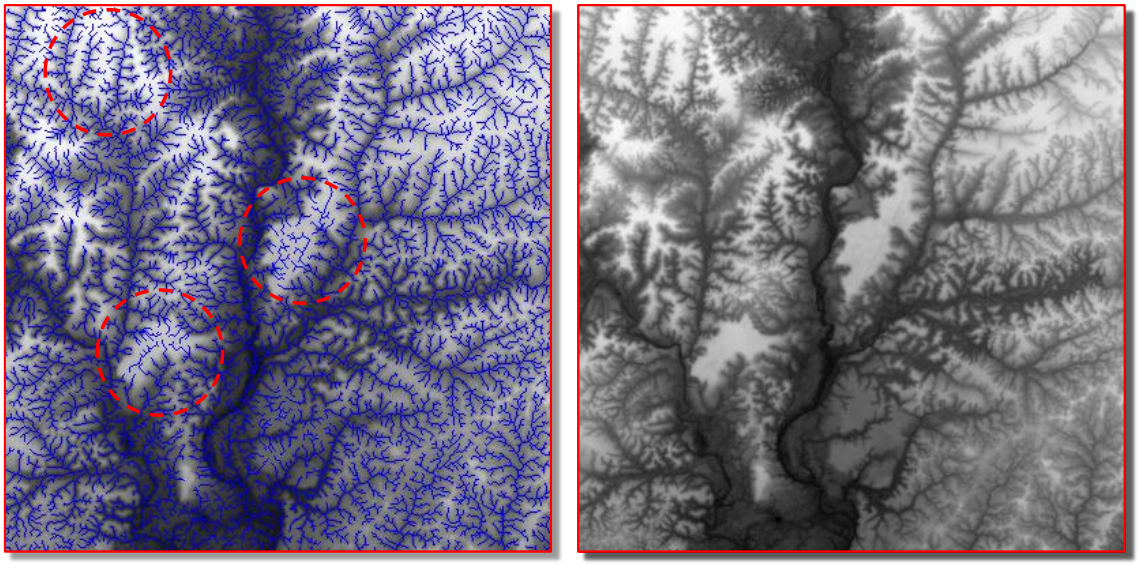


Figura 4.5 – Rede de drenagem (em azul) sobre o MDE SRTM, obtida a partir do primeiro conjunto amostral da metodologia proposta.

Na etapa de avaliação, os testes comparativos apontaram que o resultado desta primeira iteração do processo de mineração apresentou um ganho significativo em relação à metodologia clássica (BANON, 2013). Das 50 comparações realizadas, 68% (valor $p = 0{,}0077$) indicaram uma melhora ao usar a metodologia proposta.

Conforme definido na metodologia proposta, enquanto constatou-se um ganho significativo entre as redes de drenagem anterior e atual, foram realizadas novas iterações no fluxo de mineração de dados.

4.2.2. Fase Final: Seleção das Amostras, Árvore de Decisão e Avaliação

A cada iteração, novos pontos foram escolhidos para as 3 classes. Após algumas iterações obteve-se o resultado final. O conjunto amostral desta última iteração é apresentado na Figura 4.6, totalizando 615 pontos, sendo 354 da classe não drenagem, 89 da classe nascente e 172 da classe drenagem. Note que a proporção inicial do número de amostras para cada classe foi alterada durante as iterações.

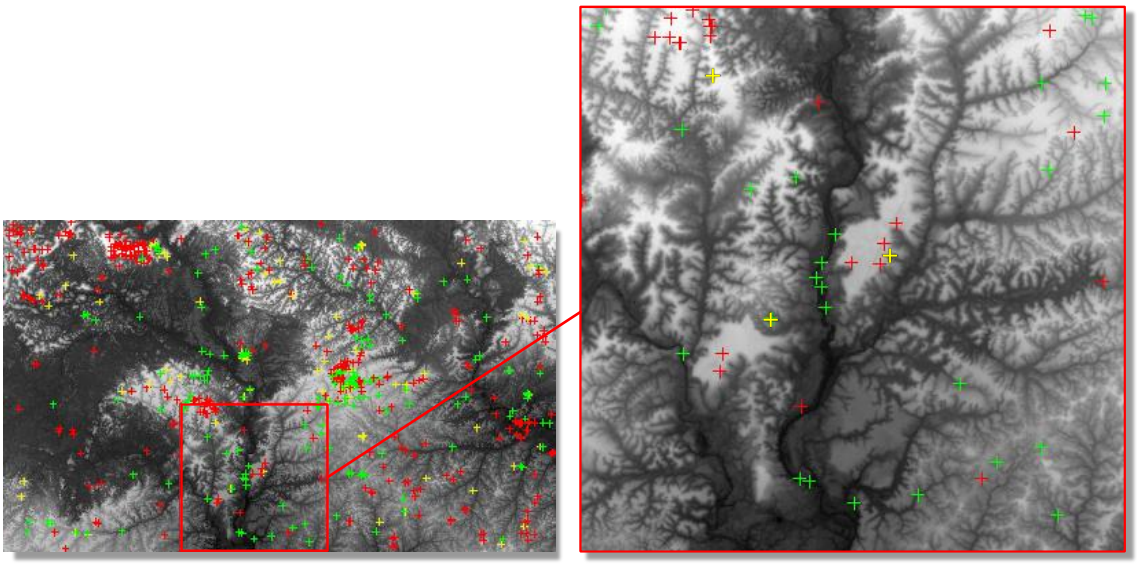


Figura 4.6 – Conjunto amostral resultante da última etapa de seleção das amostras após a fase de revisão dos pontos.

No decorrer das iterações, conforme o conjunto amostral foi sendo ampliado, as árvores resultantes apresentavam mais níveis e atributos, configurando uma estrutura mais complexa em relação às obtidas nas iterações iniciais. Com alguns ajustes nos parâmetros do J48, foi possível simplificar a estrutura da árvore de decisão. Desta forma, durante as iterações foram realizados testes e, dos valores padrões indicados pelo WEKA para o algoritmo J48, adotou-se uma mudança apenas nos parâmetros Confidence factor e MinNumObj.

O parâmetro MinNumObj (identificado no WEKA pela letra M) define o número mínimo de amostras por folha na árvore de decisão. De acordo com os testes

realizados, foi escolhido o valor 5 para este parâmetro, que inicialmente oferece o 2 como valor padrão. Ao aumentar o valor deste atributo, cada folha que antes continha apenas 2 amostras agora precisaria ter no mínimo 5. Nesta circunstância, o classificador faz uma reestruturação da árvore, unindo os elementos de duas ou mais folhas para que cada folha resultante tenha pelo menos 5 amostras. O resultado é uma poda da árvore original, ou seja, ocorre uma simplificação. Como este processo resulta muitas vezes na união de folhas de classes distintas, há um aumento esperado nos erros de classificação. A revisão e o reposicionamento destes pontos incorretos diminuíram as inconsistências, pois em geral estes erros eram casos particulares ou pontos sobre feições no limite da definição entre duas classes. Sendo assim, observou-se que o ajuste deste parâmetro permitiu a simplificação da estrutura da árvore de decisão e a diminuição das redundâncias no processo de classificação.

O parâmetro confidenceFactor (identificado no WEKA pela letra C) define o fator de confiança que é usado no processo de poda da árvore de decisão. Diminuindo este fator, aumenta-se a poda e o número de erros por folha. Neste trabalho, observou-se que em alguns casos, a diminuição do fator 0,25 para 0,1 simplificou a árvore, sem alterar o resultado da classificação. A Figura 4.7 apresenta um comparativo entre os valores 0,25 e 0,1, ao serem utilizados para a geração da árvore de decisão da última iteração da metodologia proposta. A Árvore 1, gerada a partir do valor 0,25, possui as folhas F9, F10, F11 e F12 (em destaque) correspondendo às classes de nascente ou drenagem. Nesta mesma figura, a Árvore 2 gerada com o valor 0,1, apresenta uma estrutura bem mais simples em relação à anterior, pois as amostras das quatro folhas em destaque na Árvore 1 foram condensadas em uma única folha na Árvore 2 (identificada como F9). Uma vez que as nascentes pertencem à rede de drenagem, esta simplificação apresentou o mesmo resultado de classificação. Sendo assim, neste trabalho adotou-se a estrutura da Árvore 2 como a árvore de decisão Final.

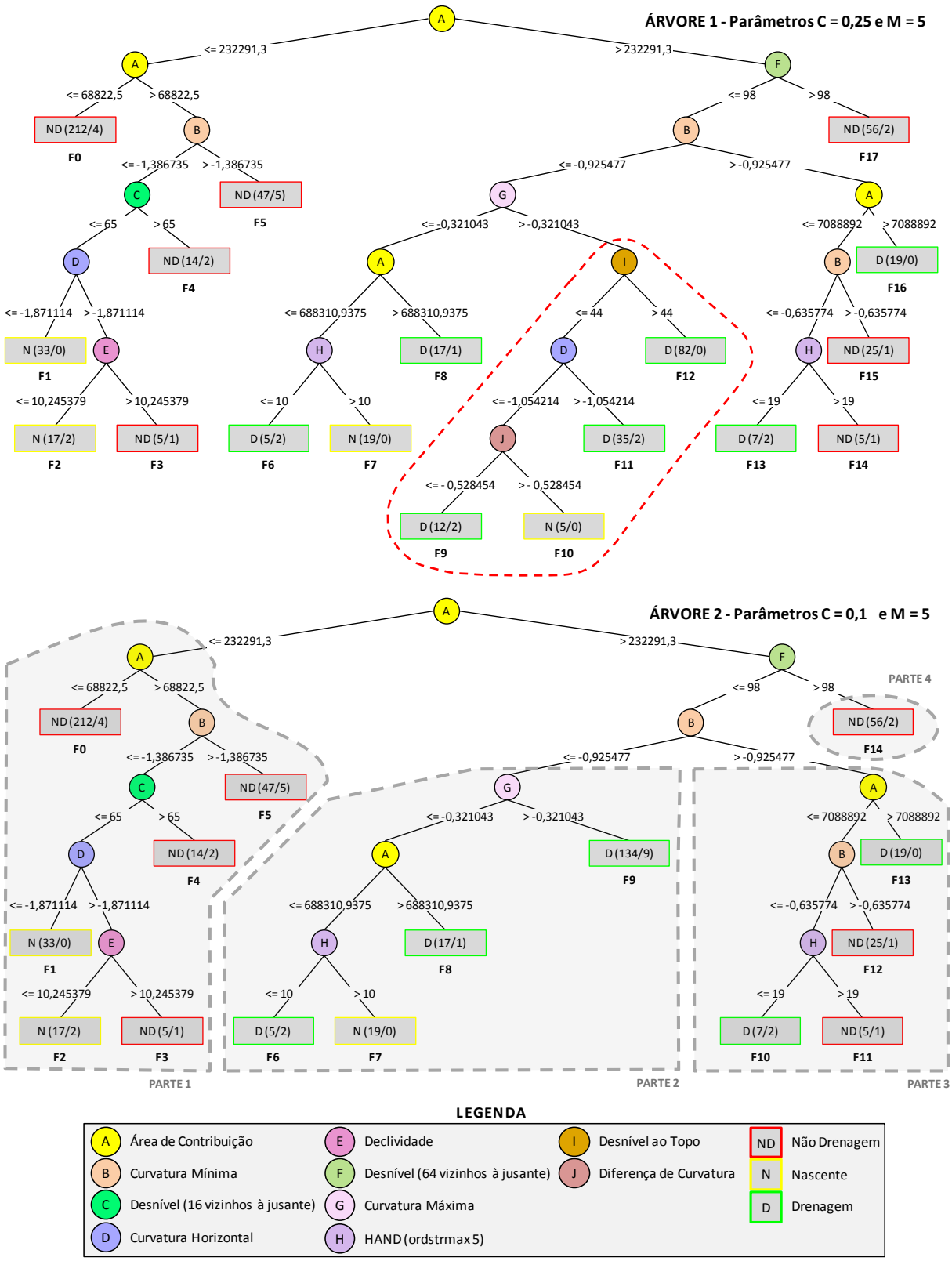


Figura 4.7 – Comparativo entre as árvores geradas pelo J48, usando o Fator de Confiança de 0,25 e 0,1. As folhas F0 a F17 possuem entre parênteses o número total de amostras classificadas e os erros de classificação.

Na Figura 4.7, é possível observar que a Árvore 2 pode ser dividida em quatro partes. A primeira parte abrange os casos que compreendem os menores valores de área de contribuição. Esta parte é responsável pela diferenciação entre nascente e não drenagem. Em geral, o limite entre estas duas classes é muito tênue.

Grande parte da classe não drenagem (folha F0) está associada a baixos valores de área de contribuição, menores que 68.822 m^2. Este limiar é muito menor do que o empregado na metodologia clássica que associa a classe não drenagem a regiões com área de contribuição menor que 435.000 m^2. A metodologia proposta permite que a rede de drenagem se inicie no intervalo entre estes dois limiares, adotando valores intermediários de área de contribuição. Neste intervalo, encontram-se os casos de nascente e não drenagem. Para diferenciar estas duas classes, a metodologia proposta utiliza alguns atributos como as curvaturas (kmin e horcurv), a declividade e o difz16. O kmin (curvatura mínima) e o difz16 (desnível de um ponto em relação a 16 vizinhos à jusante) possuem um papel importante, pois definem um limiar a partir do qual não devem ocorrer as nascentes. Isso significa que estes atributos causam um efeito de poda na rede de drenagem, impedindo o prolongamento de falsos trechos à montante. Esta poda mais acentuada se deve ao limiar do kmin, enquanto o difz é responsável pela poda mais sutil no início das cabeceiras. A Figura 4.8 apresenta a sobreposição da classificação referente às folhas F4 e F5 da Árvore 2 sobre o MDE SRTM. Note o grande número de falsos trechos de drenagem que são eliminados pela presença destas duas folhas. Desta forma, o efeito indesejável do avanço de falsos trechos de drenagem à montante é neutralizado pela marcação dos pontos de não drenagem referentes às folhas F4 e F5.

O atributo horcurv representa a curvatura horizontal. Os valores negativos deste atributo indicam áreas de convergência, que geralmente estão associadas às nascentes. Na árvore, é possível identificar que valores muito negativos (que indicam alta convergência) estão de fato relacionados às

nascentes (folha F1). Nos casos em que a convergência não é tão acentuada, cabe à declividade definir a diferenciação entre as classes de nascente (folha F2).e não drenagem (folha F3). Neste caso, as nascentes estão associadas às regiões menos declivosas.

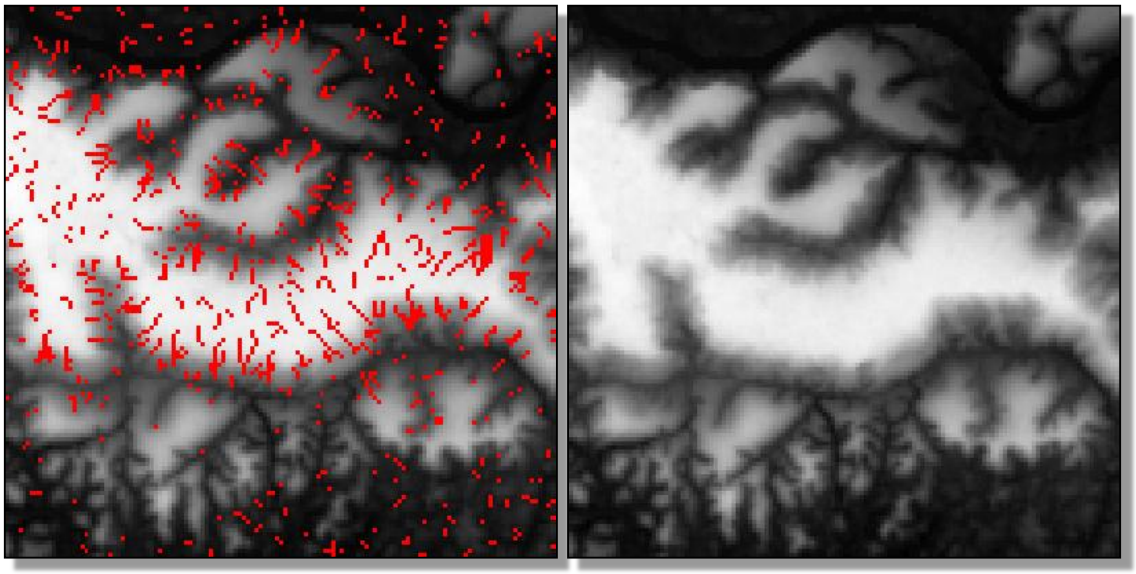


Figura 4.8 – Pontos em vermelho da classe não drenagem, pertencentes as folhas F4 e F5 da Árvore 2, impedem o avanço da rede de drenagem sobre o platô (áreas claras do MDE SRTM).

Na segunda e terceira parte da árvore, é possível observar que os valores mais altos de área de contribuição estão geralmente relacionados à rede de drenagem. A segunda parte tenta diferenciar as classes de nascente e drenagem, mas esta diferenciação não é relevante na definição da rede de drenagem, pois ambas a representam. A terceira parte é mais complexa, pois trata da diferenciação entre drenagem e não drenagem para áreas de contribuição mais elevadas. Neste caso, novamente a curvatura mínima e a área de contribuição aparecem como atributos importantes, incluindo também o atributo handordstrmax5 (folhas F10 e F11). Valores baixos para o atributo hand estão associados a regiões onde o lençol freático está próximo a superfície, identificando regiões próximas a grandes rios, cujos valores do atributo ordstrmax são elevados.

Na quarta parte da árvore, há um caso particular em que a área de contribuição também é alta, mas um elevado valor para o atributo difz64 diferencia a classe não drenagem do restante do grupo (folha F14). Neste caso, a árvore está indicando que um ponto muito distante em relação a um grande rio, ao possuir um desnível muito elevado, terá uma grande chance de estar associado à ocorrência de uma não drenagem.

A partir da aplicação da Árvore 2 sobre o conjunto amostral, obteve-se a matriz de confusão apresentada na Tabela 4.1. A confusão entre as classes nascente e drenagem deve ser desconsiderada da análise da matriz de confusão, pois ambas pertencem à rede de drenagem. Desta forma, efetivamente foram observadas poucas confusões, com apenas 22 erros de classificação no total.

Tabela 4.1 – Matriz de confusão obtida na Fase Final da metodologia proposta.

		CLASSIFICAÇÃO		
		Não Drenagem	Drenagem	Nascente
REF	Não Drenagem	**348**	2	4
	Drenagem	12	**67**	10
	Nascente	4	0	**168**

A aplicação da Árvore 2 apresentou resultados bastante promissores quando comparados à metodologia clássica (Figura 4.9). A grande contribuição da metodologia proposta pode ser observada tanto nas regiões com padrões dendríticos como nas áreas associadas aos platôs. No círculo mais acima da Figura 4.9a, é possível observar que a rede de drenagem está subestimada e nos círculos inferiores, falsos trechos de drenagem avançam sobre o platô. Em contrapartida, na Figura 4.9b estas mesmas regiões apresentam redes de drenagem classificadas adequadamente.

Os testes comparativos apontaram que a metodologia proposta apresentou resultados com um ganho muito significativo (BANON, 2013). Nesta avaliação, todas as 50 comparações realizadas indicaram que a metodologia proposta era

superior à metodologia clássica, o que pode ser comprovado na Figura 4.10 que exemplifica o resultado de duas avaliações realizadas.

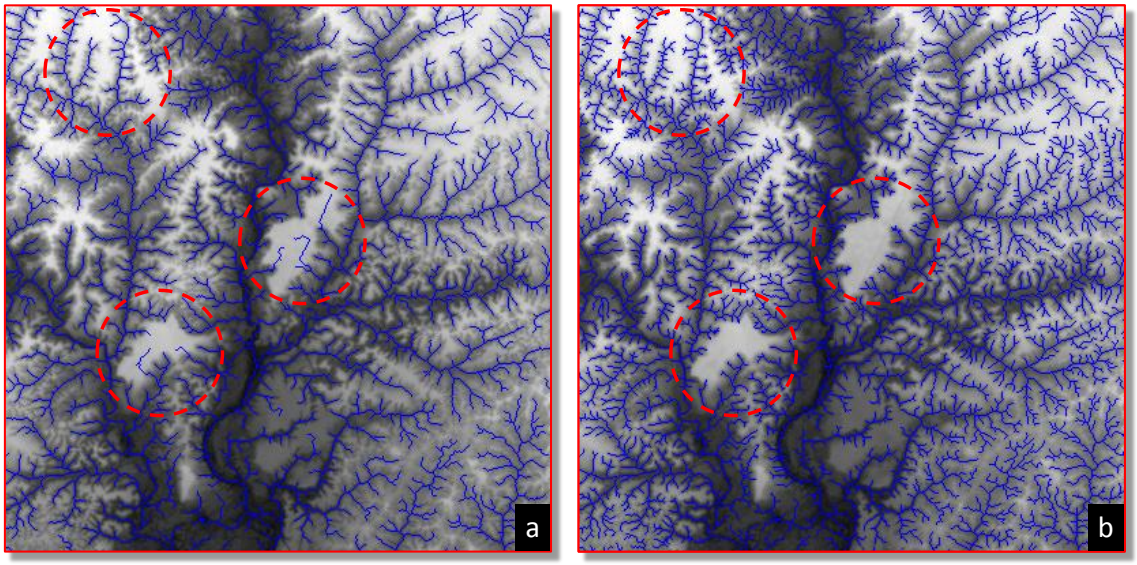


Figura 4.9 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) metodologia clássica e b) metodologia proposta. Os círculos destacam as áreas cujos resultados são mais discrepantes.

A comparação entre as metodologias clássica e proposta também foi realizada a partir do cruzamento entre as redes de drenagem resultantes destas duas metodologias. Na Figura 4.11, é apresentado um exemplo deste teste comparativo, na qual é possível observar, em azul, os trechos de drenagem comuns às duas metodologias; em ciano, aqueles pertencentes exclusivamente à rede obtida pela metodologia proposta; e, em vermelho, os referentes somente à metodologia clássica. Note que no canto superior esquerdo desta figura, a região com padrão dendrítico possui pequenos e numerosos trechos de drenagem que foram classificados apenas pela metodologia proposta. No centro desta figura, há uma rede de drenagem com feições mais tênues, onde é possível observar em vermelho a presença de falsos trechos de drenagem gerados pela metodologia clássica. Em contrapartida, nesta mesma região, notam-se várias extremidades e pequenos trechos em ciano, comprovando que a metodologia proposta permitiu a classificação adequada dos trechos com os padrões mais sutis.

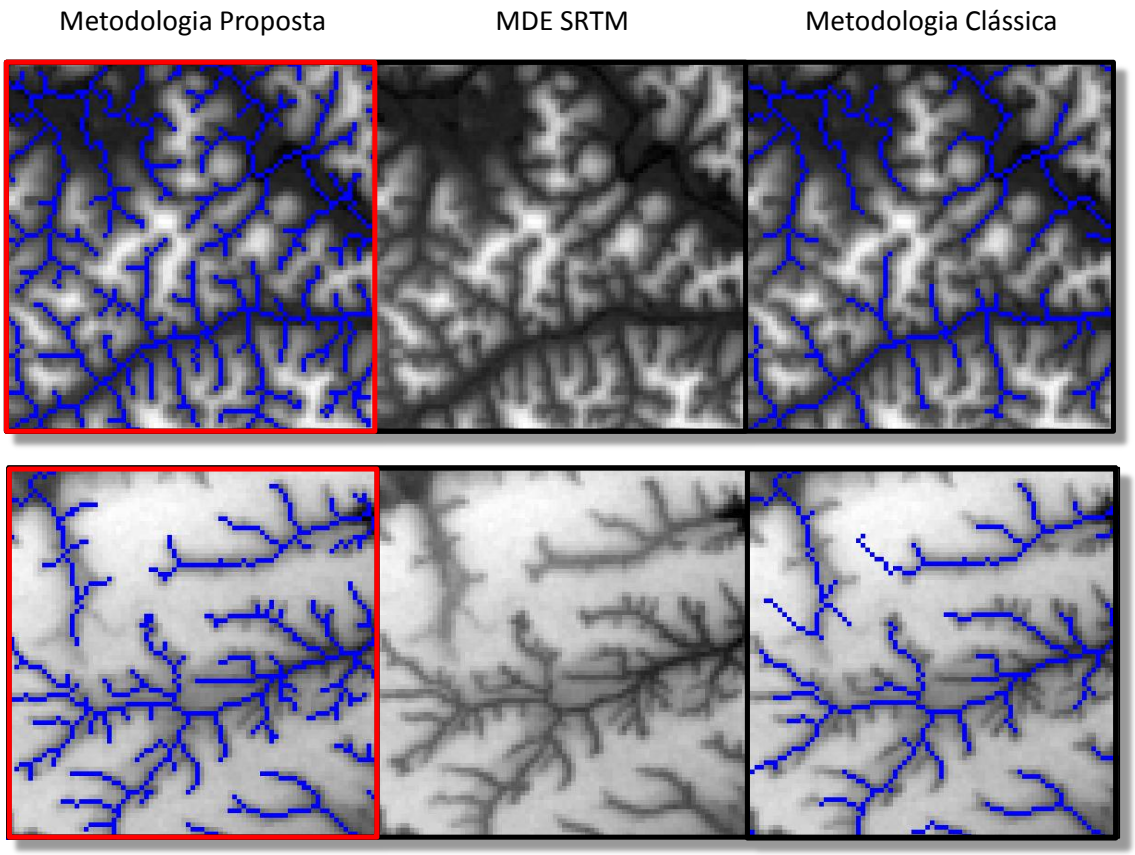


Figura 4.10 – Exemplos de duas avaliações com a seleção em vermelho do melhor resultado.

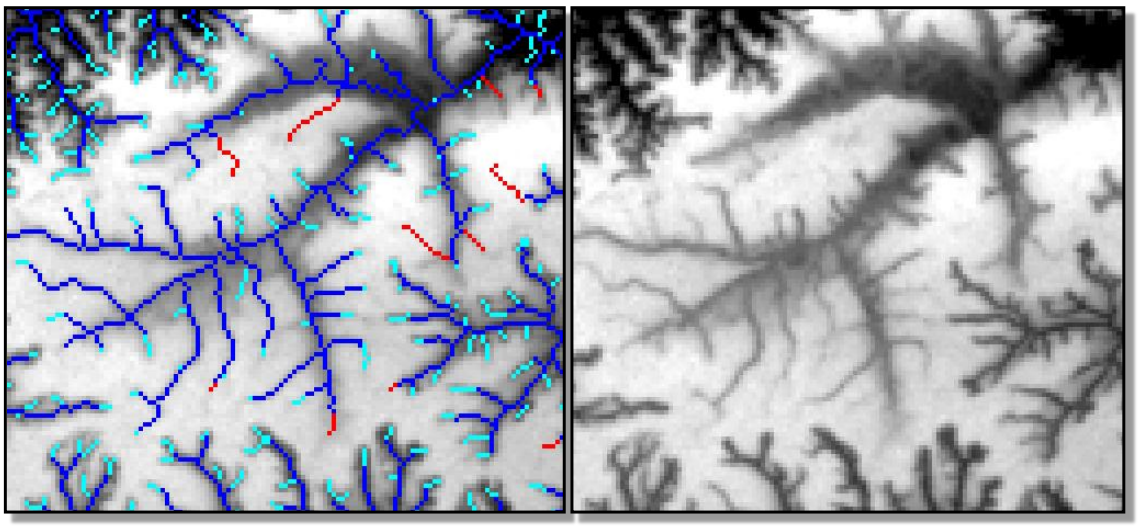


Figura 4.11 – Cruzamento das redes de drenagem obtidas a partir da metodologia clássica e proposta sobre o MDE SRTM. Em azul são os trechos comuns às duas redes, em ciano os exclusivos da metodologia proposta e em vermelho os exclusivos da metodologia clássica.

4.2.3. Testes de Simplificação

4.2.3.1. Simplificação do número de classes

A metodologia proposta utiliza três classes: nascente, drenagem e não drenagem. Para verificar a viabilidade de considerar somente duas classes (drenagem e não drenagem), sem comprometer significativamente os resultados, realizou-se um teste no qual a única modificação foi incorporar as amostras da classe nascente à classe drenagem. A partir do novo conjunto amostral, foi obtida a matriz de confusão e a árvore de decisão. Em seguida, os resultados da classificação gerados pelo novo conjunto amostral, foram comparados aos obtidos pela metodologia proposta.

Inicialmente, com base na matriz de confusão (Tabela 4.2), verificou-se que o número total de erros de classificação é 22, o mesmo da metodologia proposta. Por sua vez, a árvore de decisão (Figura 4.12) gerada a partir desta abordagem apresentou-se com uma estrutura bem mais simples com apenas 9 folhas, enquanto a obtida pela metodologia proposta apresentou 15 folhas. Analisando as folhas da árvore obtida nesta abordagem, observou-se que a folha F4, que representa a classe drenagem, foi a mais numerosa. No entanto, nesta folha encontra-se maior número de erros de classificação. Como os erros nesta folha de drenagem são da classe de não drenagem, o impacto na grade será a expansão da rede de drenagem. Assim como na metodologia proposta, os atributos área de contribuição (areacontr) e curvatura mínima (kmin) apresentam-se como os dois atributos principais.

Tabela 4.2 – Matriz de confusão obtida pela abordagem com a simplificação do número de classes.

		CLASSIFICAÇÃO	
		Não Drenagem	Drenagem
REF	Não Drenagem	**342**	12
	Drenagem	10	**251**

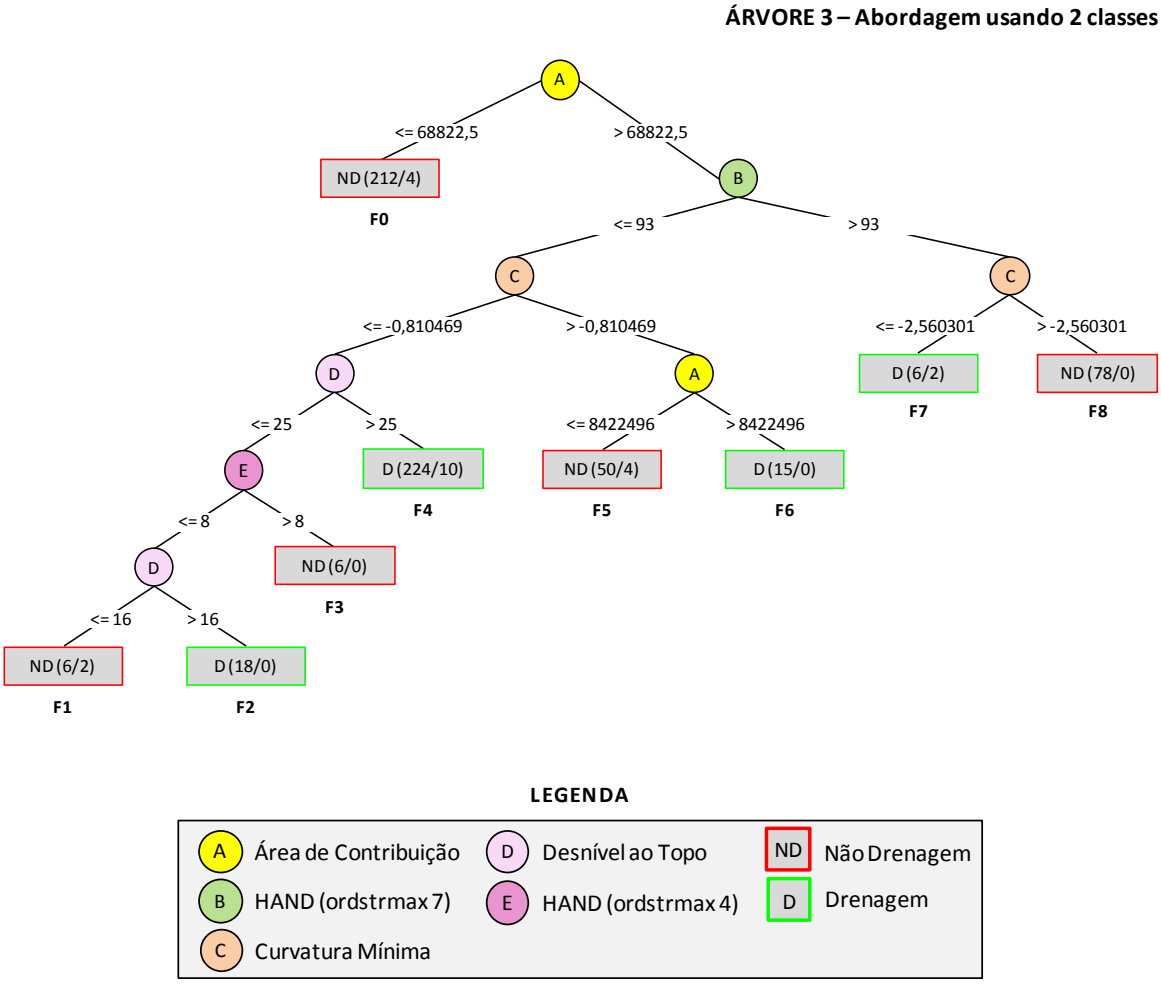


Figura 4.12 – Árvore de decisão 3 utilizando apenas duas classes. Para as folhas F0 a F8, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação.

Na etapa de avaliação, realizou-se a comparação a partir do cruzamento entre as redes de drenagem resultantes das abordagens adotando 2 e 3 classes. Um exemplo deste teste comparativo é apresentado na Figura 4.13, na qual é possível observar que a abordagem com 2 classes detectou drenagens mais curtas e numerosas, enquanto que a utilização de 3 classes permitiu classificar drenagens extensas, em regiões mais tênues. A supressão destas drenagens mais extensas tem um impacto mais relevante para a rede de drenagem do que a perda de numerosos pequenos trechos.

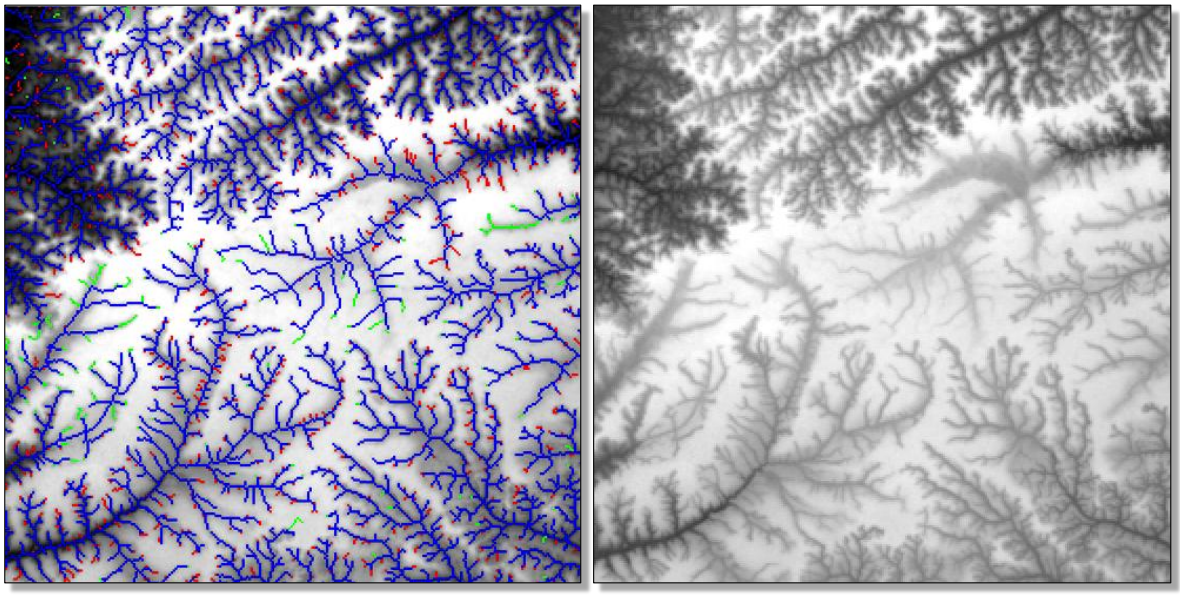


Figura 4.13 – Cruzamento das redes de drenagem obtidas a partir da utilização de 2 e 3 classes sobre o MDE SRTM. Em azul são representados os trechos comuns às duas redes, em verde os exclusivos para a abordagem usando 3 classes e em vermelho os exclusivos para 2 classes.

Outro aspecto importante observado em relação à abordagem com 2 classes, refere-se ao fato dos trechos curtos e numerosos causarem um adensamento excessivo de regiões próximas às nascentes, localizadas adjacentes aos platôs. O início de uma rede de drenagem por definição é tênue e, portanto este adensamento das nascentes é uma inconsistência que deve ser levada em consideração. A Figura 4.14 apresenta o resultado de duas análises realizadas na etapa de comparação das duas abordagens, para a avaliação do ganho significativo. Nesta figura é possível observar o adensamento das nascentes destacadas em vermelho nas áreas circulares. Nesta etapa de avaliação, os testes comparativos apontaram que a utilização de 3 classes apresentou um ganho significativo em relação à utilização de 2 classes (BANON, 2013). Das 50 comparações realizadas, 9 não apresentaram diferença e 73% das restantes foram melhor avaliadas ao usar as 3 classes, ou seja, observou-se uma melhora significativa (valor-p = 0,0022) adotando-se 3 classes ao invés de 2.

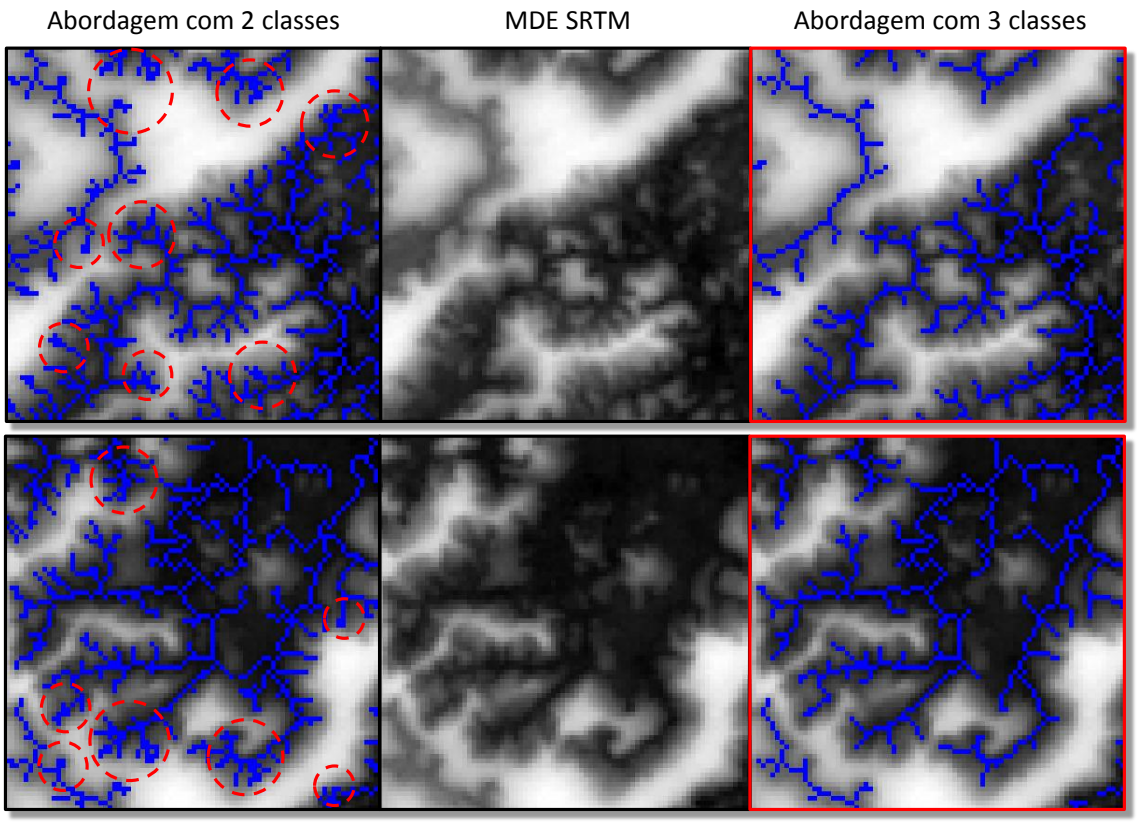


Figura 4.14 – Exemplos de duas avaliações com o melhor resultado selecionado com um retângulo vermelho. Os círculos destacam as áreas mais discrepantes.

4.2.3.2. Simplificação usando apenas atributos morfométricos

Empregando apenas os atributos morfométricos e o conjunto amostral da última etapa da metodologia proposta, foi obtida uma árvore de decisão. Em seguida, os resultados desta classificação selecionando apenas os atributos morfométricos foram comparados aos obtidos pela metodologia proposta.

Inicialmente, avaliou-se a matriz de confusão obtida pela simplificação da metodologia proposta usando apenas atributos morfométricos (Tabela 4.3). Observou-se uma confusão considerável entre as classes drenagem e não drenagem, totalizando 51 erros, enquanto as classes não drenagem e nascente apresentaram 23 erros.

Tabela 4.3 – Matriz de confusão obtida pela abordagem usando apenas atributos morfométricos.

		CLASSIFICAÇÃO		
		Não Drenagem	Drenagem	Nascente
REF	Não Drenagem	**329**	21	4
	Drenagem	30	**138**	4
	Nascente	19	20	**50**

Na Figura 4.15, é apresentada a Árvore 4 gerada exclusivamente a partir dos atributos morfométricos. Dos 18 atributos utilizados nesta abordagem de simplificação, a Árvore 4 selecionou 11 sendo que os atributos declividade e curvatura plana (plancurv) foram utilizados mais de uma vez. Ao contrário da árvore obtida na metodologia proposta, esta árvore apresenta uma estrutura extremamente complexa e a interpretação do seu significado hidrológico não é trivial.

Para uma melhor avaliação da estrutura da Árvore 4, foi realizada uma análise observando o impacto da classificação de cada folha desta árvore no MDE SRTM. Neste processo, a folha F1 foi detectada como a responsável pela maior parte dos erros de classificação, resultando em muitos pontos dispersos e não representativos sobre a grade. Esta folha representa 69% das células classificadas como drenagem. Na Árvore 4, é possível observar que a folha F1 possui 33 erros, dos quais 14 são amostras de não drenagem, representando portanto 10,5% desta folha (Figura 4.16). Uma folha com a inclusão de muitos erros de não drenagem provoca uma expansão da rede, pois significa que a classificação está considerando muitos casos de não drenagem como sendo drenagem.

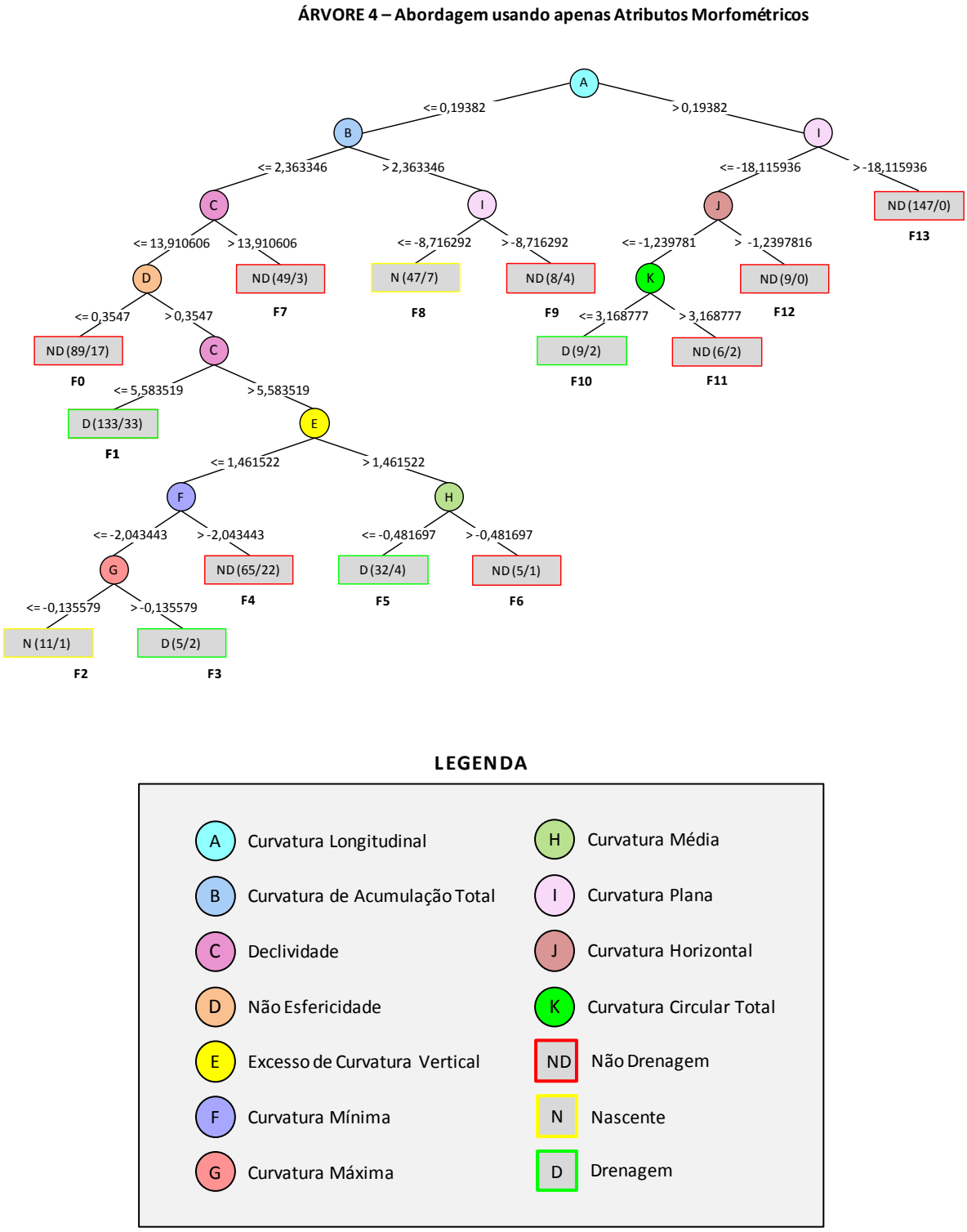


Figura 4.15 – Árvore de decisão 4 utilizando apenas atributos morfométricos. Para as folhas F0 a F13, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação.

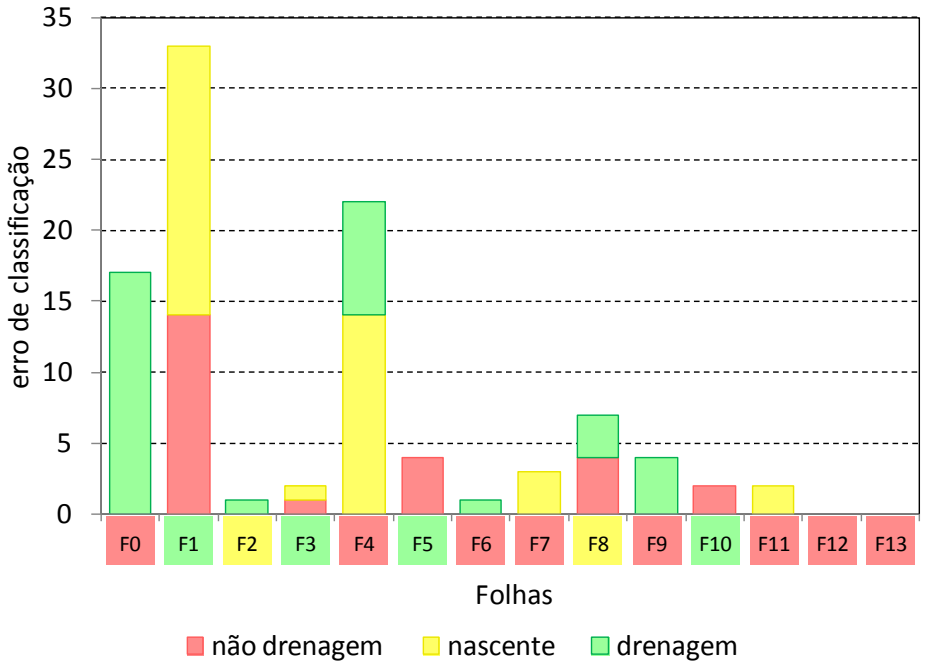


Figura 4.16 – Gráfico indicando o número de erros por folha da Árvore de decisão 4.

Na etapa de avaliação, os testes comparativos apontaram que o resultado da metodologia proposta apresentou um ganho muito significativo em relação a sua simplificação usando apenas os atributos morfométricos (BANON, 2013). Todas as 50 comparações realizadas foram melhores ao utilizar todos os atributos, ou seja, observou-se uma melhora muito significativa (valor-p < 0,00001) ao usar a metodologia proposta.

A aplicação da simplificação da metodologia proposta usando apenas atributos morfométricos apresentou resultados bastante insatisfatórios (Figura 4.17). Conforme observado na análise da árvore de decisão desta abordagem, a classificação incorreta de várias amostras de não drenagem ocasionou uma expansão indevida da rede de drenagem, com a criação de muitas feições inexistentes. A Figura 4.18 apresenta o resultado de duas análises realizadas na etapa de comparação das duas abordagens, para a avaliação do ganho significativo. Nesta figura é possível observar o avanço generalizado da rede de drenagem.

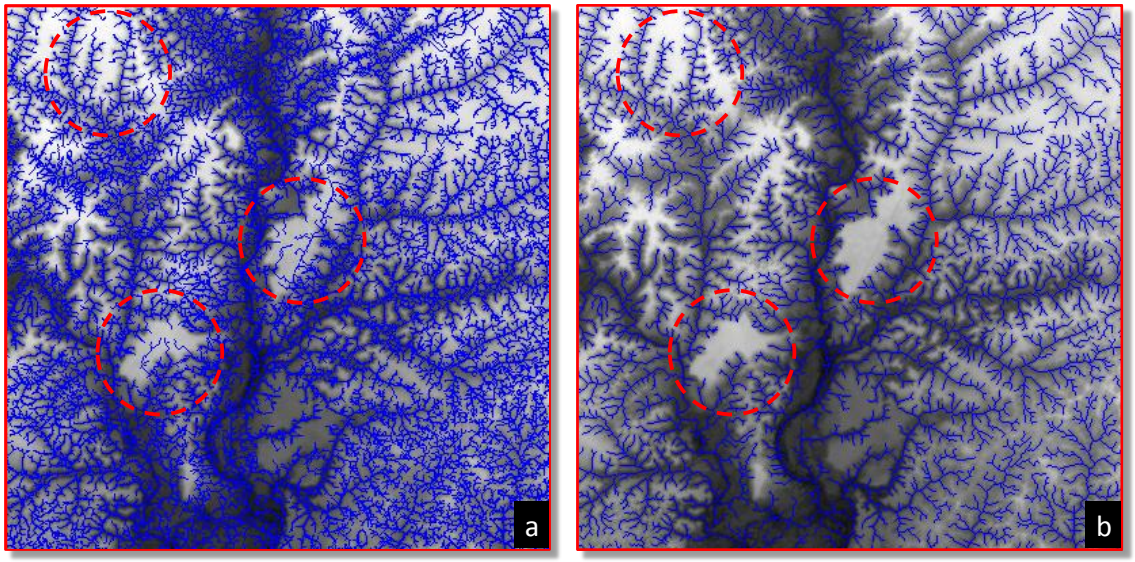


Figura 4.17 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) abordagem usando apenas atributos morfométricos e b) metodologia proposta. Os círculos destacam algumas áreas divergentes.

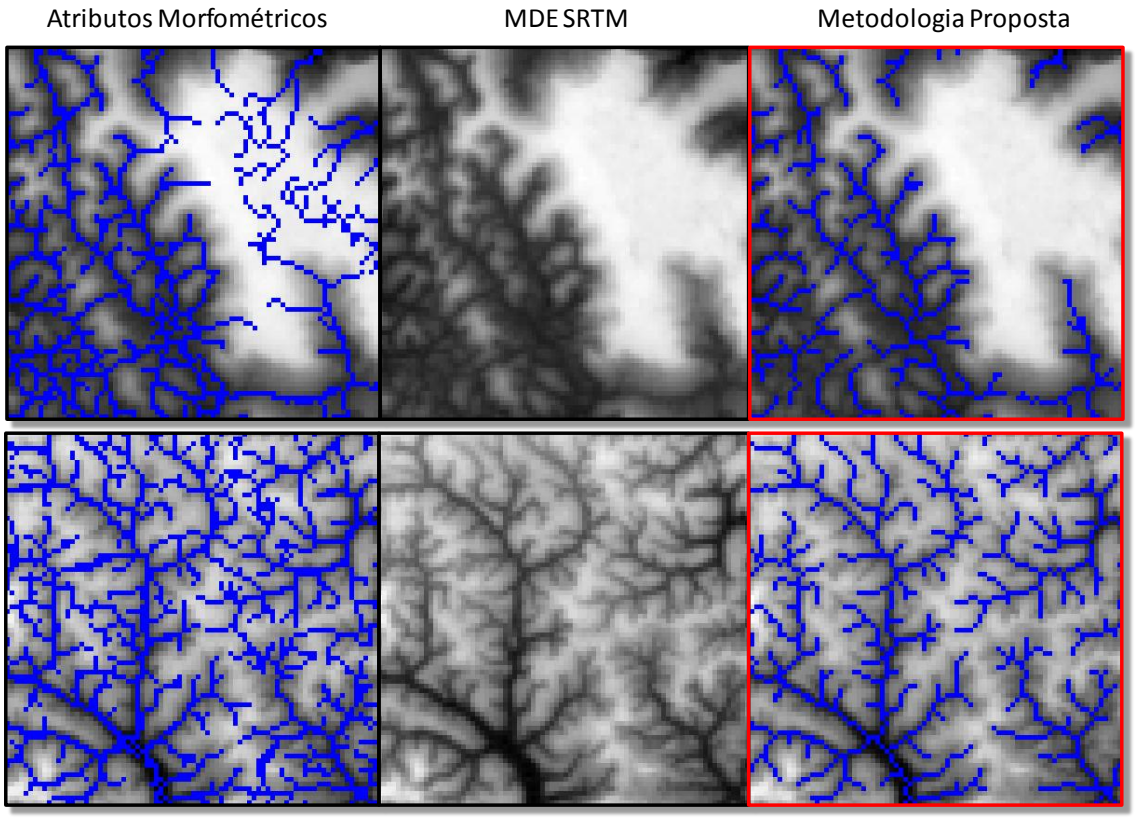


Figura 4.18 – Exemplos de duas avaliações com o melhor resultado selecionado com um retângulo vermelho.

4.2.3.3. Simplificação usando apenas a área de contribuição

A partir do conjunto amostral da última etapa da metodologia proposta, as amostras de nascente e drenagem foram agrupadas em uma única classe. Em seguida, utilizando o classificador J48 sobre este conjunto amostral reformulado e apenas o atributo área de contribuição, obteve-se a matriz de confusão (Tabela 4.4) e a árvore de decisão.

Tabela 4.4 – Matriz de confusão obtida pela abordagem simplificada usando apenas a área de contribuição.

		CLASSIFICAÇÃO	
		Não Drenagem	Drenagem
REF	Não Drenagem	**208**	146
	Drenagem	4	**257**

A avaliação da matriz de confusão obtida pela abordagem simplificada indicou muitos erros de classificação, totalizando 150 casos de confusão entre as 2 classes consideradas. Este elevado número era esperado, uma vez que a área de estudo é heterogênea e esta abordagem utiliza um único atributo.

A partir da árvore de decisão (Figura 4.19) é possível observar que na abordagem simplificada foi definido o valor 68.822,5 m^2 para o limiar da área de contribuição. Este valor é bem inferior ao limiar adotado pela metodologia clássica de 435.000 m^2. Sendo assim, espera-se que a rede de drenagem desta abordagem seja mais densa quando comparada à da metodologia clássica.

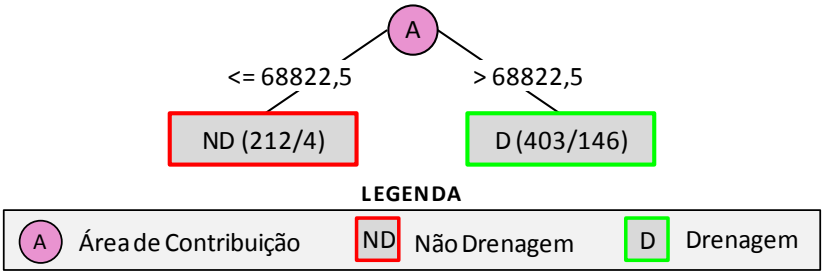


Figura 4.19 – Árvore de decisão obtida pela abordagem usando apenas a área de contribuição. Para cada folha, os números entre parênteses representam, respectivamente, o total de amostras classificadas e os erros de classificação.

Nos testes comparativos para a avaliação do ganho significativo, observou-se que das 50 comparações realizadas, houve cinco empates e, para 51,1% do restante, ajustar o valor do limiar da área de contribuição para um conjunto de treinamento foi melhor do que adotar o limiar da metodologia clássica (BANON, 2013). Desta forma, não se observou um ganho significativo ao utilizar a abordagem simplificada (valor-p = 0,5).

Na Figura 4.20, é possível comparar o resultado da classificação da abordagem simplificada usando apenas o atributo área de contribuição (Figura 4.20a) em relação à metodologia clássica (Figura 4.20b). Note que nas regiões com padrões dendríticos, como a área exemplificada no círculo mais acima, a abordagem simplificada resultou em uma classificação adequada, enquanto que a metodologia clássica resultou em uma rede de drenagem subestimada. Por outro lado, nas áreas associadas aos platôs, exemplificadas nos dois círculos mais inferiores, a abordagem simplificada criou várias feições inexistentes, enquanto a metodologia clássica apresentou o mesmo resultado, mas em número menor.

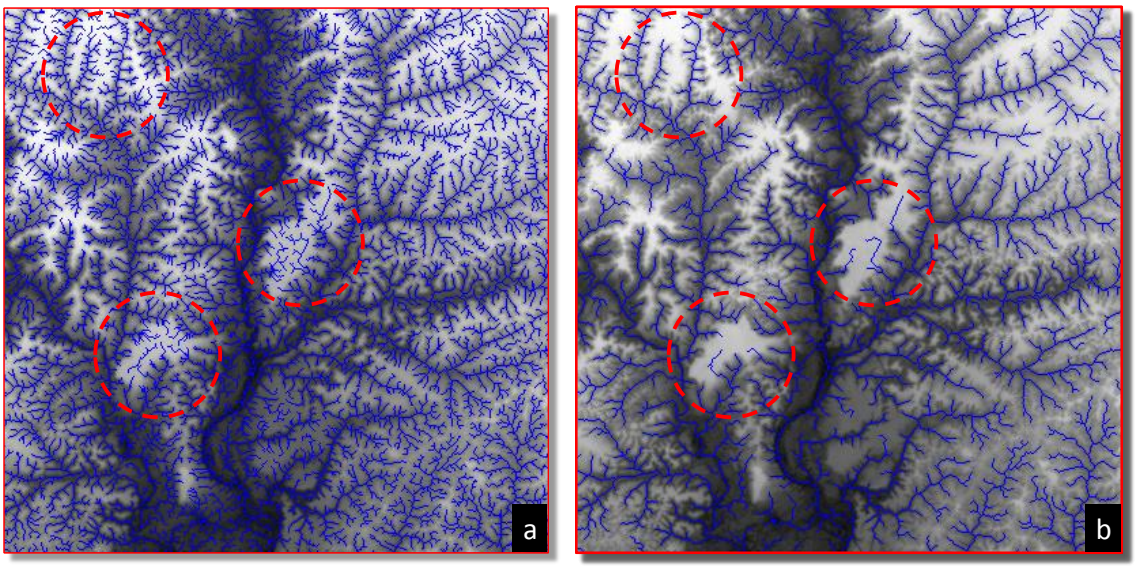


Figura 4.20 – Comparação entre a rede de drenagem (em azul) sobreposta ao MDE SRTM, gerada pela: a) abordagem usando apenas a área de contribuição e b) metodologia clássica. Os círculos destacam algumas áreas divergentes.

Considerando que não houve uma melhora significativa da abordagem simplificada em relação à metodologia clássica, tornou-se desnecessário realizar a comparação entre a abordagem simplificada e a metodologia proposta.

4.2.4. Testes de Robustez

Seguindo a metodologia proposta obteve-se uma árvore de decisão na área escolhida para treinamento (Figura 3.1a). Em seguida, esta árvore foi aplicada em quatro áreas teste na região amazônica (Figuras 3.1b, 3.1c, 3.1d e 3.1e). O intuito desta abordagem foi verificar a robustez desta árvore de decisão, ou seja, avaliar se a mesma era extrapolável para outras regiões, de modo a obter resultados superiores à metodologia clássica.

Inicialmente, ao aplicar a árvore de decisão avaliou-se o impacto de cada uma de suas folhas na grade das áreas teste. A representatividade de cada folha indica a porcentagem de células da rede de drenagem que foram classificadas como pertencentes a esta folha. A representatividade relativa às folhas das classes nascente e drenagem é detalhada na Tabela 4.5.

Tabela 4.5– Representatividade das folhas pertencentes as classes de nascente e drenagem da árvore de decisão da metodologia proposta na área de treinamento e nas áreas teste.

FOLHA	CLASSE	ÁREA A (treinamento)	ÁREA B (teste)	ÁREA C (teste)	ÁREA D (teste)	ÁREA E (teste)
F1	nascente	7,5%	0,0%	4,2%	0,3%	0,0%
F2	nascente	11,1%	1,1%	11,7%	3,8%	0,6%
F6	drenagem	2,7%	0,8%	2,6%	1,6%	0,7%
F7	nascente	2,2%	0,1%	1,6%	0,2%	0,1%
F8	drenagem	5,8%	2,2%	5,6%	5,6%	3,0%
F9	drenagem	52,2%	36,4%	56,1%	47,3%	36,4%
F10	drenagem	8,7%	31,9%	10,5%	21,6%	33,7%
F13	drenagem	9,9%	27,4%	7,7%	19,5%	25,5%

É importante notar que a áreas B e E possuem um padrão de distribuição de representatividade bem distintos da área de treinamento. Particularmente, a

folha F10 consiste em apenas 8,7% da rede de drenagem da área de treinamento, enquanto nas áreas B e E, esta folha representa mais de 30% da classificação da rede de drenagem (valores destacados em vermelho na Tabela 4.5). Isso ocorreu porque na área de treinamento a folha F10 era responsável pela classificação de algumas feições raras e em contrapartida, nas áreas B e E havia uma maior ocorrência destas feições. Esta observação evidencia a importância da escolha de uma área de treinamento com uma grande diversidade de padrões geomorfológicos para a aplicação da metodologia proposta.

A folha F10 classificou feições importantes nas áreas teste, mas em alguns casos também resultou na classificação de pontos isolados e distantes da rede de drenagem. A Figura 4.21 exemplifica o impacto da classificação da folha F10 sobre a grade do MDE SRTM numa região da área B. Note que várias feições desta folha estão corretamente associadas à rede de drenagem. No entanto, alguns pontos isolados são responsáveis pelo aparecimento de várias feições inexistentes.

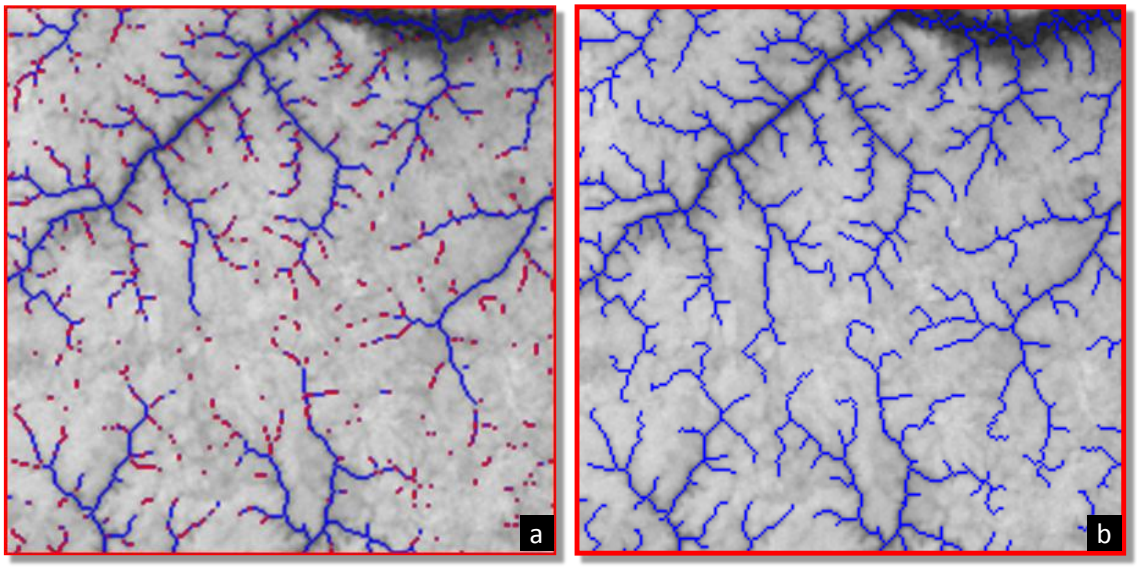


Figura 4.21 – Rede de drenagem sobreposta ao MDE SRTM da área B. a) com as células referentes a folha F10 destacadas em vermelho e a demais células da rede de drenagem em azul e b) rede de drenagem (em azul) resultante da conexão de todos os trechos isolados.

Observou-se que em geral, estes casos ocorreram devido à limitação dos dados do MDE SRTM. Estes dados podem sofrer a influência do uso e cobertura, que pode ser natural ou de origem antrópica. Nestes casos, a área apresenta um padrão indefinido, prejudicando o reconhecimento das feições. A Figura 4.22 apresenta o exemplo de uma área antropizada da Área B, na qual se torna difícil definir o início da rede de drenagem.

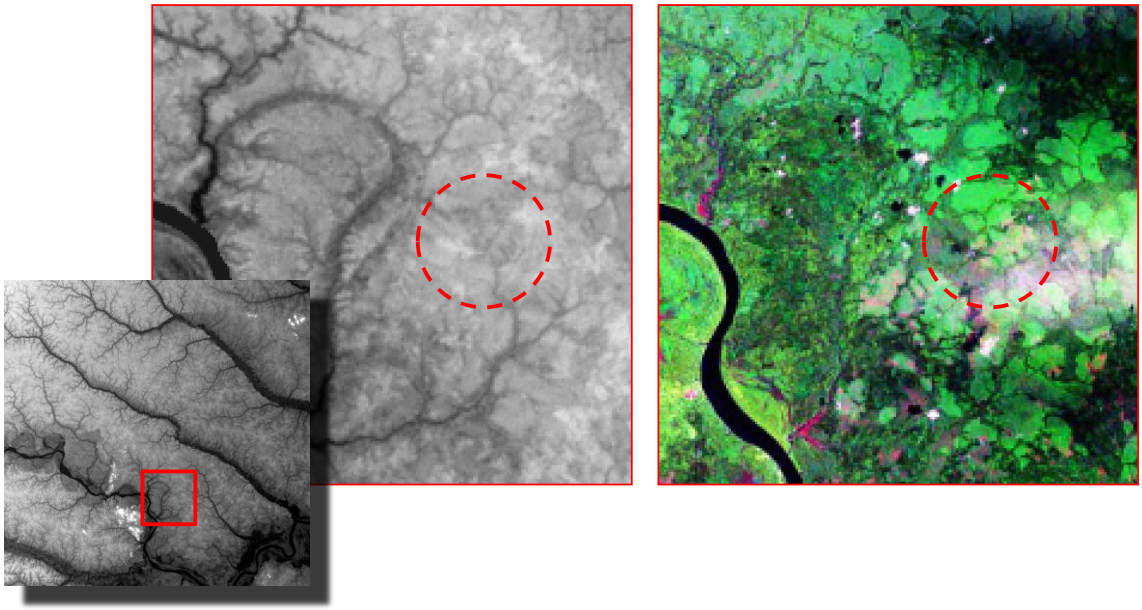


Figura 4.22 – MDE SRTM da Área B com área circular destacando uma região antropizada, dificultando o reconhecimento das nascentes. Ao lado, a imagem Geocover indicando esta influência (áreas magentas).

Ao utilizar o pós-processamento convencional, que corrige todas as descontinuidades da rede de drenagem (utilizando os parâmetros *gapsize* e *segsize* iguais a zero), observou-se que a presença de pontos isolados e distantes, resultou em uma inadequada expansão desta rede. Desta forma, definiu-se uma etapa de pós-processamento com ajuste fino da rede de drenagem, a partir da escolha dos melhores parâmetros. Neste trabalho, apenas trechos cujo intervalo era inferior ou igual a 5 células foram conectados (*gapsize* = 5). Em seguida, todos os trechos que permaneceram isolados e que possuíam tamanho inferior ou igual a 3 células foram apagados (*segsize* = 3). Por fim, os trechos remanescentes e isolados, foram conectados.

A Figura 4.23 apresenta um comparativo entre a metodologia clássica; a aplicação da árvore da metodologia proposta na Área B com pós-processamento convencional (*gapsize* e *segsize* iguais a zero) e esta mesma técnica com o pós-processamento utilizando um ajuste fino (*gapsize* = 5 e *segsize* = 3). Nota-se que a aplicação da árvore da metodologia proposta com o pós-processamento utilizando um ajuste fino, obteve resultados muito superiores em relação às demais abordagens.

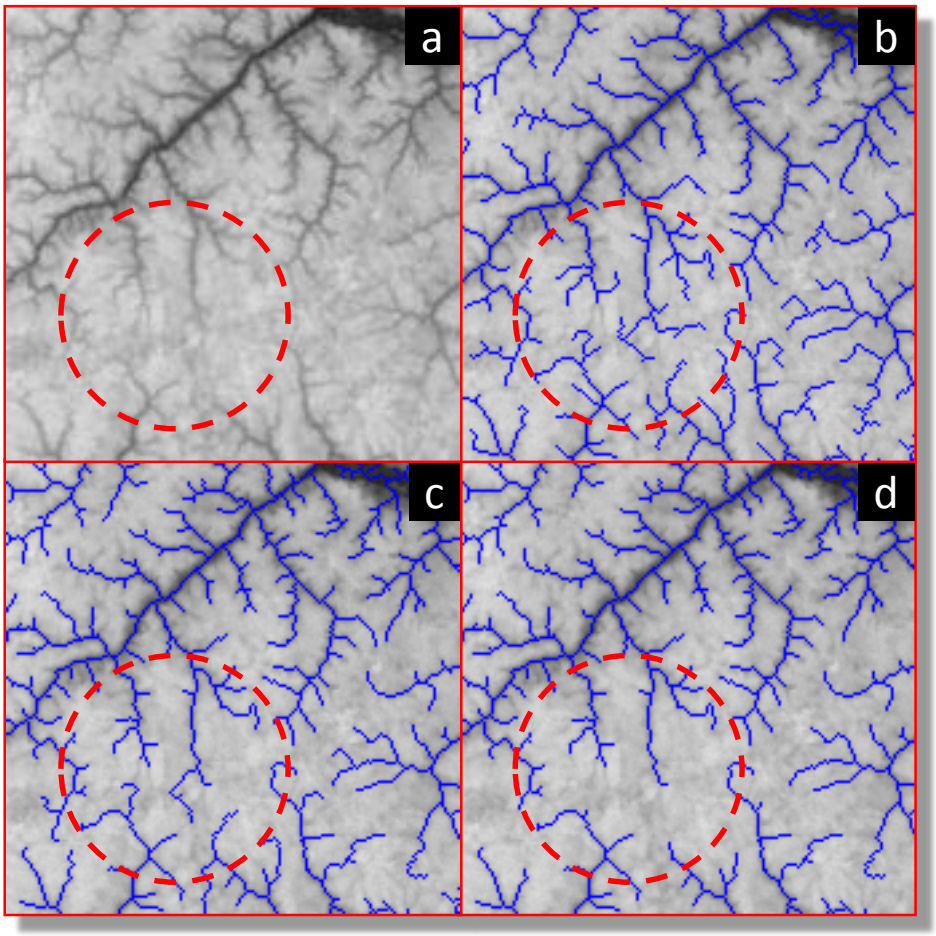


Figura 4.23 – Comparação entre os resultados das metodologias clássica e proposta numa quadrícula da área B. a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem obtida pela aplicação da árvore da metodologia proposta usando *gapsize* e *segsize* iguais a zero e d) rede de drenagem obtida pela aplicação da árvore da metodologia proposta usando *gapsize* = 5 e *segsize* = 3. O círculo destaca uma área discrepante.

Na etapa de avaliação, foram realizados os testes comparativos para a análise do ganho significativo. Verificou-se que os resultados da metodologia proposta com ajuste fino apresentou ganho significativo em relação à metodologia proposta com pós-processamento convencional, que por sua vez apresentou um resultado superior em relação à metodologia clássica (BANON, 2013).

A Tabela 4.6 apresenta o resultado da avaliação do ganho significativo para todas as áreas teste. Em todas as avaliações, observou-se que o pós-processamento adotando *gapsize* = 5 e *segsize* = 3 sobre a aplicação da árvore da metodologia proposta obteve os melhores resultados.

Tabela 4.6– Resultado das 50 Avaliações do Ganho Significativo.

Área	Comparação*	Empates	p	valor-p
B	AMP x MC	17	0,848	< 0,0001
	AMPp x AMP	11	0,744	0,0017
C	AMP x MC	1	0,980	< 0,0001
	AMPp x AMP	28	0,909	0,0001
D	AMP x MC	1	0,980	< 0,0001
	AMPp x AMP	21	0,862	0,0001
E	AMP x MC	1	0,918	< 0,0001
	AMPp x AMP	19	0,903	< 0,0001

* MC: metodologia clássica; AMP: Aplicação da árvore da metodologia proposta com pós-processamento convencional; e AMPp: Aplicação da árvore da metodologia proposta com pós-processamento de ajuste fino.

As Figuras 4.24, 4.25, 4.26 e 4.27 apresentam a comparação entre as diferentes metodologias utilizadas para a obtenção da rede de drenagem para as quatro áreas teste B, C, D e E, respectivamente.

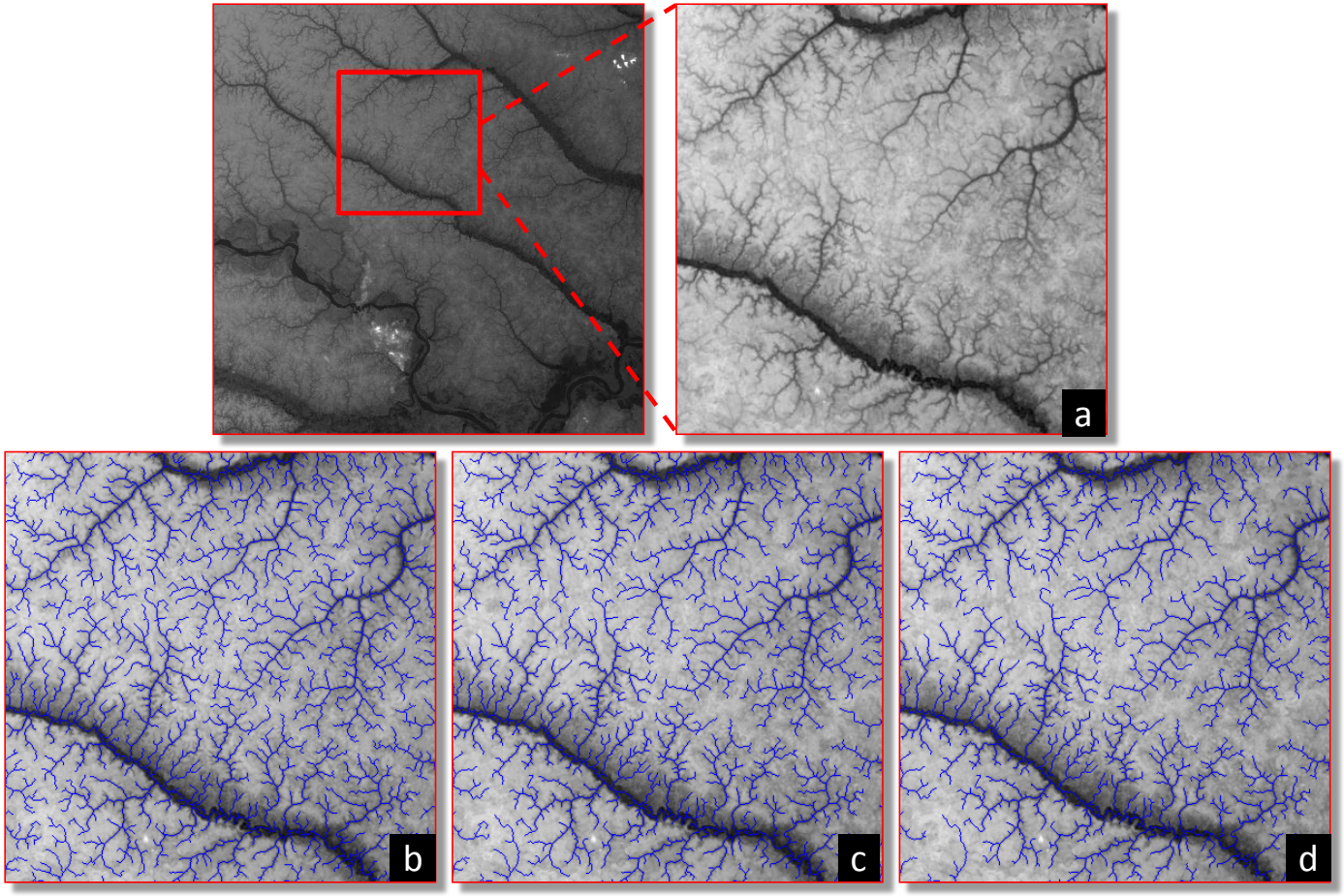


Figura 4.24 – Resultados da Área B: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino.

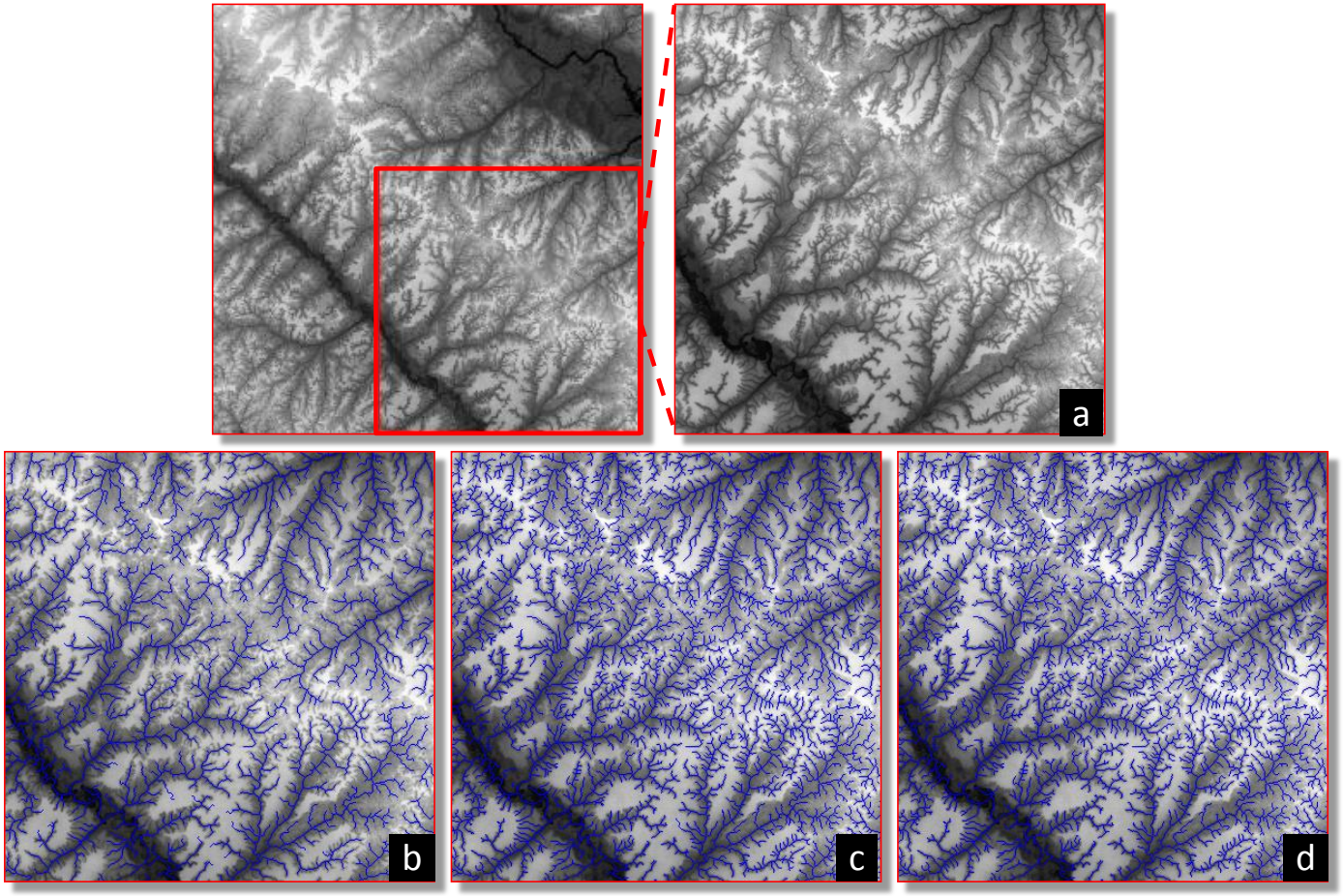


Figura 4.25 – Resultados da Área C: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino.

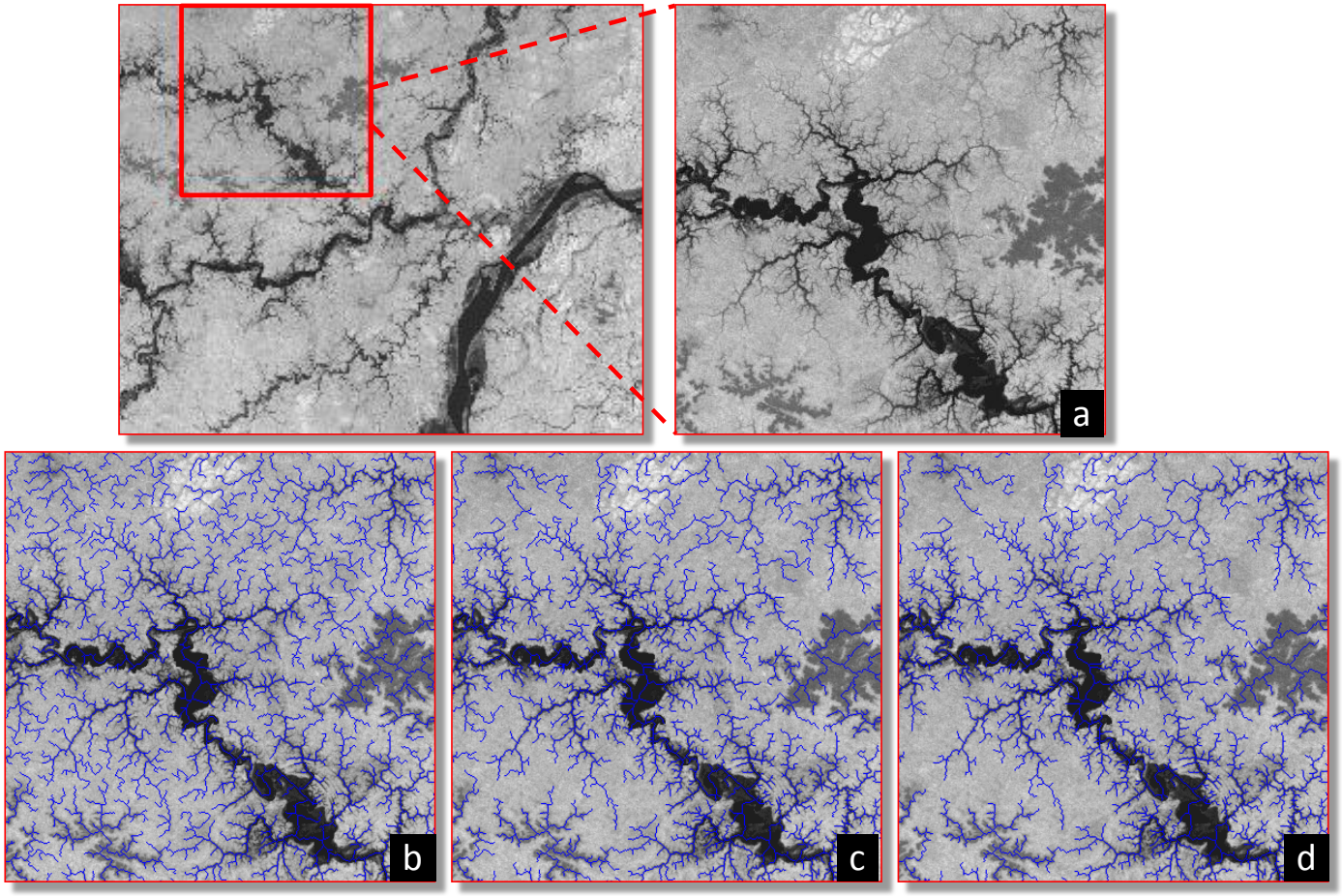


Figura 4.26 – Resultados da Área D: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino.

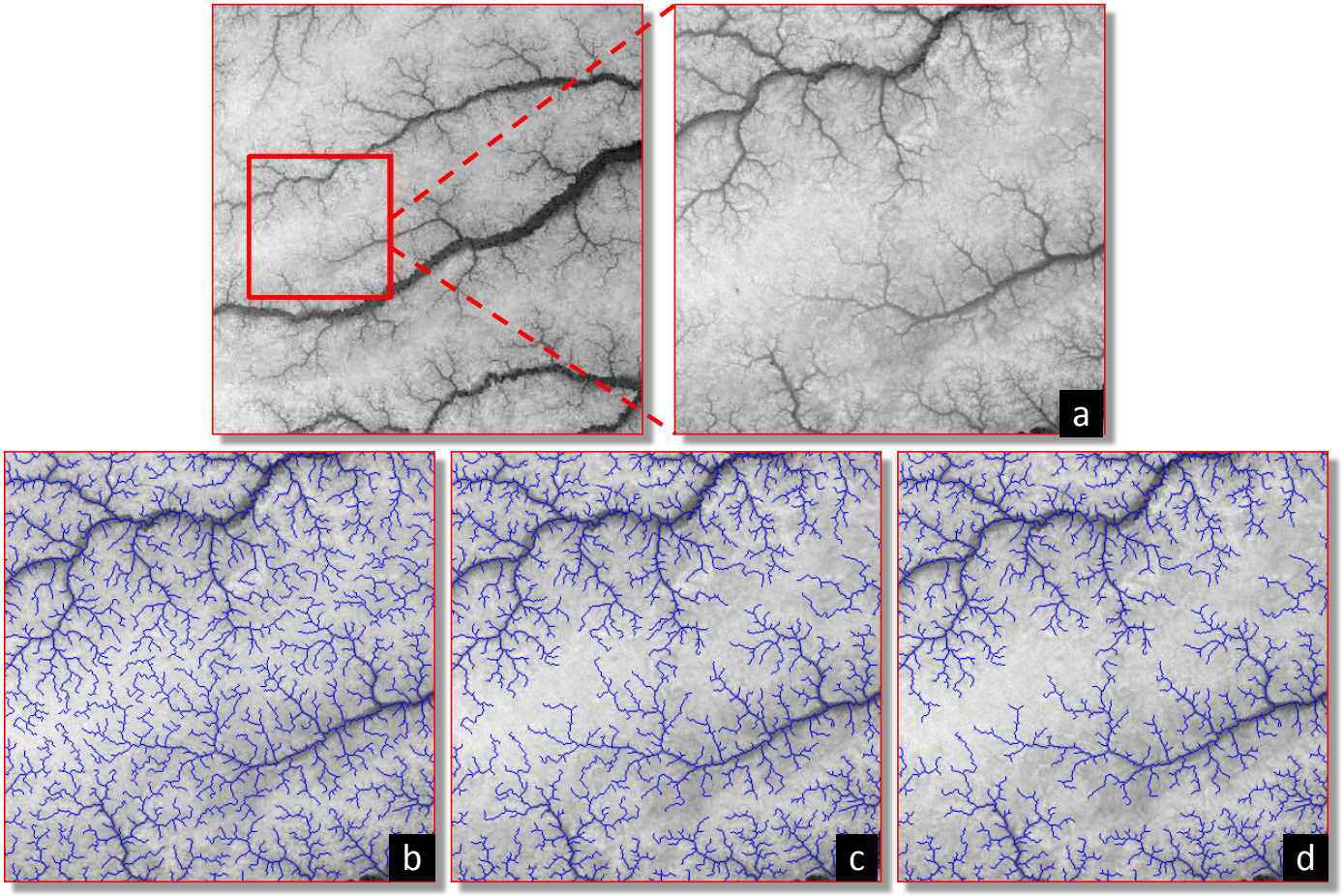


Figura 4.27 – Resultados da Área E: a) MDE SRTM; b) rede de drenagem obtida pela metodologia clássica; c) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento convencional; e d) rede de drenagem resultante da aplicação da árvore de decisão obtida pela metodologia proposta com pós-processamento de ajuste fino.

5 CONSIDERAÇÕES FINAIS

Tradicionalmente, a extração automática de rede de drenagem é baseada na escolha empírica de um limiar para um único atributo: a área de contribuição. No entanto, este atributo não é suficiente para representar a rede de drenagem em regiões com diferentes padrões geomorfológicos.

O presente trabalho apresentou uma nova metodologia para a extração automática de redes de drenagem baseada em um conjunto de atributos extraídos do MDE SRTM. Conceitualmente, estes atributos podem ser divididos em dois grupos: os morfométricos e os baseados na direção de fluxo. Esta metodologia utilizou o algoritmo J48, um classificador do tipo árvore de decisão, com o qual foi possível selecionar os atributos mais representativos e desta forma, definir critérios objetivos para a classificação da rede de drenagem.

Para aplicar a metodologia proposta, foram selecionadas cinco áreas de estudo na região amazônica. A primeira área foi escolhida por apresentar uma grande diversidade de padrões geomorfológicos, sendo selecionada para aquisição de um conjunto amostral bem representativo. A partir deste conjunto amostral obteve-se uma árvore de decisão que foi aplicada nas outras áreas em estudo a fim de testar a sua robustez.

No total foram coletados 615 pontos representando as 3 classes: nascente, drenagem e não drenagem. Com base neste conjunto amostral, obteve-se a árvore de decisão que selecionou os seguintes atributos: área de contribuição, curvaturas mínima, máxima e horizontal, desnível ao vizinho à jusante e o HAND (desnível à drenagem mais próxima). É importante salientar que outros atributos, além da área de contribuição, foram essenciais para a diferenciação entre as três classes em estudo.

A partir da árvore de decisão foi gerada a rede de drenagem da primeira área de estudo, cuja qualidade foi muito superior em relação à metodologia clássica, baseada unicamente na área de contribuição. Esta rede de drenagem classificou adequadamente regiões com alta e baixa densidade de drenagem.

Em seguida, esta árvore de decisão foi aplicada nas demais áreas de estudo, gerando redes de drenagem também com qualidade superior em relação à metodologia clássica, demonstrando que esta árvore pode ser extrapolável para outras regiões.

Para trabalhos futuros, é sugerida a avaliação de outras técnicas de Mineração de Dados para a extração automática de redes de drenagem. Testes preliminares indicaram que o IBK (*Instance-Based K-nearest neighbours classifier*) e as Redes Neurais são técnicas promissoras. No entanto, a metodologia proposta é iterativa e para uma investigação mais detalhada de seus potenciais, seria preciso otimizar o processo de classificação, testando diferentes parametrizações.

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, E. S.; ROSIM, S.; RENNÓ, C. D.; OLIVEIRA, J. R. F.; JARDIM, A. C.; ORTIZ, J. O.; DUTRA, L. V. Terrahidro a distributed hydrological system to delimit large basins. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2012, Munich. **Anais...** 2012. v. 1, p. 300-301. DVD. Disponível em: <http://www.igarss2012.org/>. Acesso em: 20 fev. 2013.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. **Machine Learning**, v. 6, p. 37-66, 1991.

BANON, L. C.; SANTOS, R. D. C.; VIJAYKUMAR, N. L.; RENNÓ, C. D. Definição de critérios a partir da mineração de dados para a extração automática de redes de drenagem. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 16. (SBSR), 2013, Foz do Iguaçu. **Anais...** São José dos Campos: INPE, 2013. p. 5753-5760. DVD, Internet. ISBN 978-85-17-00066-9 (Internet), 978-85-17-00065-2 (DVD). Disponível em: <http://urlib.net/3ERPFQRTRW34M/3E7GCTQ>. Acesso em: 15 out. 2013.

BANON, L. C. **Dados utilizados no teste comparativo para avaliação do ganho significativo da metodologia proposta**. São José dos Campos: INPE, 2013.

BANON, L. C.; RENNÓ, C. D.; **DrainageDataMining**. 1.0. São José dos Campos: INPE, 2013. Disponível em: < http://urlib.net/sid.inpe.br/mtc-m19/2013/11.19.18.43>. Acesso em: 19 nov. 2013.

BEALE R.; JACKSON, T. **Neural computing**: an introduction. Cambridge: MIT Press, 1990.

BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A.; STONE, C.J. **Classification and regression trees**. Monterey: Wadsworth and Brooks, 1984. 358p.

COLLINS, S.H.; MOON, G.C. Algorithms for dense digital terrain models. **Photogrammetric Engineering and Remote Sensing**, v. 47, n. 1, p. 71-76, 1981.

CORREIA, A. H. Metodologias e resultados preliminares do projeto radiografia da Amazônia. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15. (SBSR), 2011, Curitiba. **Anais...** São José dos Campos: INPE, 2011. p. 8083-8090. DVD, Internet. ISBN 978-85-17-00056-0 (Internet), 978-85-17-00057-7 (DVD). Disponível em: <http://urlib.net/3ERPFQRTRW/39UGKLH>. Acesso em: 02 jan. 2013.

DE'ATH, G.; FABRICIUS, K. E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. **Ecology**, v. 81, n. 11, p. 3178−3192, 2000.

DIRETORIA DE SERVIÇO GEOGRÁFICO (DSG). A Diretoria de Serviço Geográfico e o Mapeamento da Amazônia**. Revista Verde-Oliva Edição Especial Ciência & Tecnologia**, v. 35, n. 197, p. 13-16, 2008. Disponível em: <http://www.exercito.gov.br/c/document_library/get_file?uuid=c334ca00-1315-4672-aecf-50bffe90f495&groupId=52610>. Acesso em: 02 jan. 2013.

DUTRA, L. V.; ELMIRO, M. T.; SOARES, B. S.; MURA, J. C.; SANTOS, J. R.; FREITAS, C. C.; ARAUJO, L. S.; ALBUQUERQUE, P. C. G.; VIEIRA, P. R.; GAMA, F. F. Assessment of digital elevation models obtained in Brazilian Amazon based on P and X band airborne interferometric data. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2002, Toronto, Canada. **Proceedings...** Toronto: IEEE, 2002. (INPE-9045-PRE/4721).

EXELIS VISUAL INFORMATION SOLUTIONS. **IDL connectivity bridges**. Washington: EXELIS. 2010. 504 p. (0509IDL71CB). Disponível em: <http://www.exelisvis.com/docs/pdf/bridges.pdf >. Acesso em: 15 jan. 2013.

FAYYAD, U.; PIATESKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in knowledge discovery and data mining**. Cambridge: MIT Press, 1996. 560 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **IBGE traz informações do solo e dos recursos minerais a Banco de Dados da Amazônia Legal**. Rio de Janeiro: IBGE, 2012. Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=2287>. Acesso em: 03 jan. 2013.

MOURA, P.; CORREIA, A. H. Utilização de Imagens Interferométricas SAR na banda X para estimativa da cota da superfície e nivelamento de massas d´água no projeto radiografia da Amazônia. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15. (SBSR), 2011, Curitiba. **Anais...** São José dos Campos: INPE, 2011. p. 8272-8279. DVD, Internet. ISBN 978-85-17-00056-0 (Internet), 978-85-17-00057-7 (DVD). Disponível em: <http://urlib.net/3ERPFQRTRW/3A6DTJP>. Acesso em: 20 fev. 2013.

O'CALLAGHAN, J. F; MARK, D. M. The extraction of drainage networks from digital elevation data. **Computer Vision, Graphics and Image Processing**, v. 28, p. 323−344, 1984.

PINHEIRO, T. F.; RENNÓ, C. D.; ESCADA, M. I. S. Utilização de um novo algoritmo descritor de terreno para o mapeamento de ambientes de terra firme na Amazônia. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. (SBSR), 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 6069-6076. DVD, On-line. ISBN 978-85-17-00044-7. (INPE-15984-PRE/10593). Disponível em: <http://urlib.net/dpi.inpe.br/sbsr@80/2008/11.18.00.00>. Acesso em: 20 fev. 2013.

QUINLAN, R. **C4.5**: programs for machine learning. San Francisco: Morgan Kaufmann, 1993. 316p.

RENNÓ, C. D.; NOBRE, A. D.; CUARTAS, L. A.; SOARES, J. V.; HODNETT, M. G.; TOMASELLA, J.; WATERLOO, M. J. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. **Remote Sensing of Environment**, v. 112, p. 3469-3481, 2008. doi: <10.1016/j.rse.2008.03.018>.

ROSIM, S.; MONTEIRO, A. M. V.; RENNÓ, C. D.; SOUZA, R. C. M.; SOARES, J. V. TerraHidro - uma plataforma computacional para o desenvolvimento de aplicativos para a análise integrada de recursos hídricos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11. (SBSR), 2003, Belo Horizonte. **Anais...** São José dos Campos: INPE, 2003. p. 2589 - 2596. CD-ROM. ISBN 85-17-00017-X. (INPE-16167-PRE/10770). Disponível em: <http://urlib.net/ltid.inpe.br/sbsr/2002/11.17.20.15 >. Acesso em: 20 fev. 2013.

SANTOS, R. D. C. Conceitos de mineração de dados na Web. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, 15., 2009, Fortaleza. **Anais...** Fortaleza: Sociedade Brasileira de Computação, 2009. v. 1, p. 41-80. ISBN 21759669. Minicursos. Disponível em: <http://www.lac.inpe.br/~rafael.santos/Docs/WebMedia/2009/webmedia2009.pdf>. Acesso em: 02 jan. 2013. (Editado por TEIXEIRA, M. M. et al.).

SHARY, P.A.; SHARAYA, L.S.; MITUSOV, A.V. Fundamental quantitative methods of land surface analysis. **Geoderma**, v. 107 (1–2), p. 1–32, 2002.

SOILLE, P.; VOGT, J.; COLOMBO, R. Carving and adaptive drainage enforcement of grid digital elevation models. **Water Resources Research**, v. 39, n.12, p. 1366−1375, 2003.

TARBOTON, D. G.; BRAS, R. L.; RODRIGUEZ-ITURBE, I. On the extraction of channel networks from digital elevation data. **Hydrological processes**, v. 5, p. 81−100, 1991.

TARBOTON, D. G.; AMES, D. P. Advances in the mapping of flow networks from digital elevation data. In: WORLD WATER AND ENVIRONMENTAL RESOURCES CONGRESS, 2001, Orlando, Florida. **Proceedings**… American Society of Civil Engineer, 2001. p. 1-10. ISBN (print): 978-0-7844-0569-7. doi: <10.1061/40569(2001)166>.

VALERIANO, M. M.; KUPLICH, T. M.; STORINO, M.; AMARAL, B. D.; MENDES JÚNIOR, J. N.; LIMA, D. J. Modeling small watersheds in Brazilian Amazonia with shuttle radar topographic mission-90 m data. **Computers and Geosciences**, v. 32, n. 8, p. 1169-1181, Outubro 2006. (INPE-14635-PRE/9637).

VAN ZYL, J.J. The shuttle radar topography mission breakthrough in remote sensing of topography. **Acta Astronautica**, v. 48, n. 5-12, p. 559-565, 2001.

WITTEN, I.; FRANK, E. **Data mining:** pratical machine learnig tools and techniques. 2. ed. San Francisco, CA: Morgam Kaufmann Publishers, 2005. 524 p.

WOOD, J.D. **The geomorphological characterisation of digital elevation models.** 1996. Thesis - University of Leicester, UK, 1996. Disponível em: <http://www.soi.city.ac.uk/~jwo/phd>. Acesso em: 02 set. 2012.

APÊNDICE A –CÁLCULO DOS ATRIBUTOS

A.1 - Atributos Morfométricos – método de Evans-Young

Segundo o método de Evans–Young, o polinômio de segundo grau, dado por

$$z = \frac{rx^2}{2} + sxy + \frac{ty^2}{2} + px + qy + z_0$$

é ajustado por mínimos quadrados a uma subgrade 3x3, com resoluções w_x, w_y e elevações $z_1, \dots, z_9$ conforme Figura A.1.

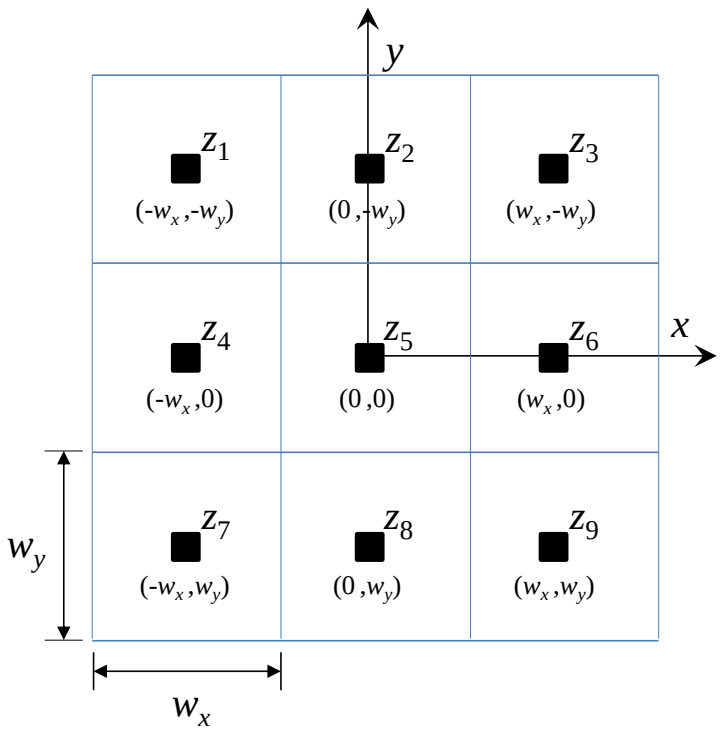


Figura A.1 – Localização e número das células da subgrade 3x3.

Os parâmetros do polinômio podem então ser estimados por

$$p = (z_3 + z_6 + z_9 - z_1 - z_4 - z_7)/6w_x$$

$$q = (z_1 + z_2 + z_3 - z_7 - z_8 - z_9)/6w_y$$

$$r = [z_1 + z_3 + z_4 + z_6 + z_7 + z_9 - 2(z_2 + z_5 + z_8)]/3w_x^2$$

$$s = (z_3 + z_7 - z_1 - z_9)/4w_x w_y$$

$$t = [z_1 + z_2 + z_3 + z_7 + z_8 + z_9 - 2(z_4 + z_5 + z_6)]/3w_y^2$$

De fato estes parâmetros representam derivadas parciais do polinômio considerando que o elemento central da subgrade tem coordenadas $x = y = 0$.

A partir da superfície ajustada, é possível calcular os seguintes atributos morfométricos:

a) Declividade (%) - slope

$$\text{slope} = 100\sqrt{p^2 + q^2}$$

b) Fator de Gradiente (100 m.m^{-1}) - gradfactor

$$\text{gradfactor} = 100\,\frac{\sqrt{p^2 + q^2}}{\sqrt{1 + p^2 + q^2}}$$

c) Curvatura Média (10^3 m^{-1}) - meancurv

$$\text{meancurv} = -1000\,\frac{(1 + p^2)t - 2pqs + (1 + q^2)r}{2(1 + p^2 + q^2)^{1,5}}$$

d) Curvatura Gaussiana Total (10^6 m^{-1}) - totgausscurv

$$\text{totgausscurv} = 10^6\,\frac{rt - s^2}{(1 + p^2 + q^2)^2}$$

e) Não Esfericidade (10^3 m^{-1}) - unsphericity

$$\text{unsphericity} = \sqrt{\text{meancurv}^2 - \text{totgausscurv}}$$

f) Curvatura Mínima (10^3 m^{-1}) - kmin

$$\text{kmin} = \text{meancurv} - \text{unsphericity}$$

g) Curvatura Máxima (10^3 m^{-1}) - kmax

$$\mathrm{kmax} = \mathrm{meancurv} + \mathrm{unsphericity}$$

h) Curvatura Vertical (10^3 m^{-1}) – vertcurv

$$\mathrm{vertcurv} = -1000\frac{p^2r + 2pqs + q^2t}{(p^2+q^2)(1+p^2+q^2)^{1,5}}$$

i) Curvatura Horizontal (10^3 m^{-1}) - horcurv

$$\mathrm{horcurv} = -1000\frac{q^2r - 2pqs + p^2t}{(p^2+q^2)(1+p^2+q^2)^{0,5}}$$

j) Diferença de Curvatura (10^3 m^{-1}) - difcurv

$$\mathrm{difcurv} = \mathrm{vertcurv} - \mathrm{meancurv}$$

k) Curvatura Longitudinal (10^3 m^{-1}) - longcurvr

$$\mathrm{longcurv} = -1000\frac{p^2r + 2pqs + q^2t}{p^2+q^2}$$

l) Curvatura Transversal (10^3 m^{-1}) - crossseccurv

$$\mathrm{crossseccurv} = -1000\frac{p^2t - 2pqs + q^2r}{p^2+q^2}$$

m) Rotor (10^3 m^{-1}) - *rotor*

$$\mathrm{rotor} = -1000\frac{(p^2+q^2)s - pq(r-t)}{(p^2+q^2)^{1.5}}$$

n) Excesso de Curvatura Vertical (10^3 m^{-1}) - kve

$$\mathrm{kve} = \mathrm{unsphericity} + \mathrm{difcurv} = \mathrm{vertcurv} - \mathrm{kmin}$$

o) Excesso de Curvatura Horizontal (10^3 m^{-1}) - khe

$$\mathrm{khe} = \mathrm{unsphericity} - \mathrm{difcurv} = \mathrm{kmax} - \mathrm{vertcurv}$$

p) Curvatura Circular Total (10^3 m^{-1}) - totringcurv

$$\mathrm{totringcurv} = \mathrm{unsphericity}^2 - \mathrm{difcurv}^2$$

q) Curvatura de Acumulação Total (10^6 m^{-1}) - totaccumcurv

$$\mathrm{totaccumcurv} = \mathrm{meancurv}^2 - \mathrm{difcurv}^2$$

r) Curvatura Plana (10^3 m^{-1}) - plancurv

$$\mathrm{plancurv} = -\frac{q^2 r - 2pqs + p^2 t}{(p^2 + q^2)^{1.5}}$$

A.2 - Atributos baseados na direção de fluxo

A direção de fluxo define a direção preferencial de escoamento da água na superfície e pode ser representada pela grade LDD. Os atributos baseados na direção de fluxo fazem uso desta grade, assim como da grade MDE. A Figura A.2 apresenta um MDE e seu respectivo LDD utilizado para explicar cada um dos atributos.

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

↘	↘	↘	↘	↘	→	·	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

Figura A.2 – Exemplo de MDE e LDD

a) Área de Contribuição - areacontr

Este atributo refere-se a área das células à montante que convergem para uma determinada célula. Este atributo pode ser expresso em número de células ou em unidades métricas. A Figura A.3 exemplifica o cálculo deste atributo. A partir de uma célula (em vermelho) são identificadas àquelas que convergem a esta (em azul). Neste exemplo, a área de contribuição está expressa em número de células.

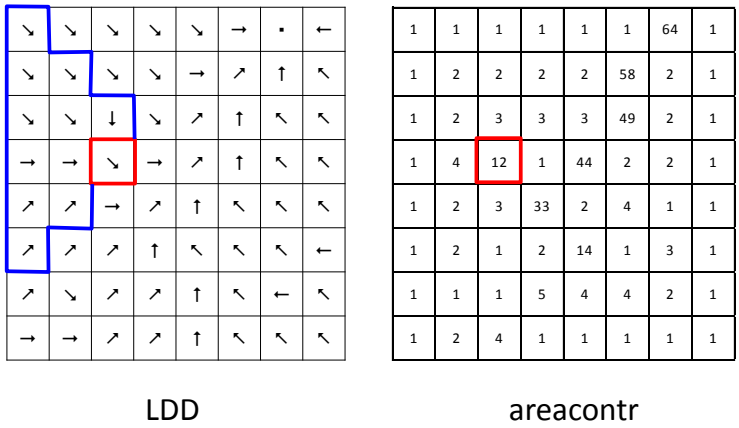


Figura A.3 – Exemplo de cálculo do atributo areacontr (em células)

b) Desnível ao Topo - desvtopo

Este atributo refere-se ao desnível máximo entre uma célula e as células à montante da mesma. A Figura A.4 exemplifica o cálculo deste atributo. A partir da grade LDD, identificam-se as células (em azul) cujos fluxos convergem para uma determinada célula (em vermelho) e verifica-se qual destas tem a maior altitude (identificado em amarelo na grade MDE). Calcula-se então a diferença entre as altitudes destas células.

↘	↘	↘	↘	↘	→	·	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

0	0	0	0	0	0	126	0
0	6	31	40	40	125	52	0
0	44	46	68	78	124	52	0
0	47	80	0	120	53	49	0
0	56	98	118	50	77	0	0
0	48	0	39	114	0	56	0
0	0	0	106	46	53	17	0
0	36	79	0	0	0	0	0

desvtopo

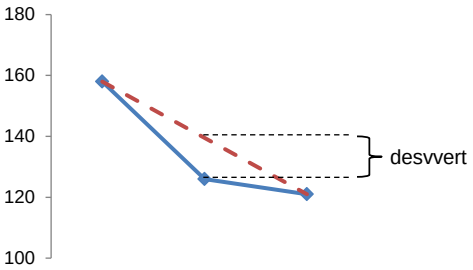
Figura A.4 – Exemplo de cálculo do atributo desvtopo

c) Desnível Vertical - desvvert

Este atributo refere-se ao desnível entre uma célula e a média altimétrica entre o máximo valor da célula à montante e a célula à jusante. A Figura A.5 exemplifica o cálculo deste atributo. A partir da grade LDD, identificam-se as células vizinhas (em azul) à montante e a célula vizinha à jusante (em verde) de uma determinada célula (em vermelho). Verifica-se qual das células à montante tem a maior altitude (identificado em amarelo na grade MDE). Calcula-se então a diferença entre a determinada célula e a média das células à montante e à jusante.

a) Declividade à jusante - decld8

Este atributo refere-se a declividade entre uma célula e sua vizinha à jusante. A Figura A.6 exemplifica o cálculo deste atributo. A partir da grade LDD, identifica-se a célula vizinha à jusante (em verde) de uma determinada célula (em vermelho). Calcula-se então a declividade (em %) entre estas células. Neste exemplo, considerou-se que a célula possuía resolução de 100m.



↘	↘	↘	↘	↘	→	·	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

0	0	0	0	0	0	-18	0
0	17	3	-1	-11	-12	-23	0
0	-6	-6	-3	-11	-24	-14	0
0	-17	-14	0	-14	-20	-0,5	0
0	-16	-18	-20	-24	4	0	0
0	1	0	-14	-19	0	-18	0
0	0	0	-9,5	-21	3	9,5	0
0	3,5	-9	0	0	0	0	0

desvvert

Figura A.5 – Exemplo de cálculo do atributo desvvert

↘	↘	↘	↘	↘	→	·	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

4,24	21,9	28,3	28,3	14,1	9,9	0	8,49
31,1	28,3	26,2	26,9	13,4	0,71	4,24	24,7
26,2	22,6	19,8	21,9	11,3	0,71	17,7	36,8
24,7	9,19	3,54	7,07	2,83	9,9	33,9	36,8
33,2	17	9,9	1,41	2,12	20,5	37,5	34,6
39,6	35,4	29	7,78	2,83	35,4	14,8	36,1
33,9	31,8	27,6	5,66	2,83	29,7	25,5	39,6
25,5	30,4	19,1	15,6	4,24	32,5	29,7	12

decld8

Figura A.6 – Exemplo de cálculo do atributo decld8

b) Ordem Máxima de Sthraler - ordstrmax

No ordenamento de Sthraler, todo segmento da rede de drenagem associado à nascente é definido como de primeira ordem. Quando dois segmentos de primeira ordem se encontram, formam um segmento de segunda ordem. Quando dois segmentos de segunda ordem se encontram formam um segmento de terceira ordem e assim sucessivamente. Neste caso, a ordem só

aumenta um nível quando dois segmentos de mesma ordem se encontram. Para o cálculo da ordem máxima de Sthraler, considera-se que todas as células da grade pertencem à rede de drenagem. A Figura A.7 exemplifica o cálculo deste atributo.

↘	↘	↘	↘	↘	→	▪	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

1	1	1	1	1	1	4	1
1	1	1	1	1	4	1	1
1	1	1	1	1	4	1	1
1	2	3	1	4	1	1	1
1	1	1	4	1	2	1	1
1	1	1	1	3	1	2	1
1	1	1	2	2	2	1	1
1	1	2	1	1	1	1	1

ordstrmax

Figura A.7 – Exemplo de cálculo do atributo ordstrmax

c) HAND com Ordem Máxima de Sthraler k – hand_ordstrmax

O atributo hand (*Height Above the Nearest Drainage*) foi definido por Rennó et al. (2008) e refere-se ao desnível de uma célula em relação a drenagem mais próxima. Neste trabalho, a rede de drenagem a qual este atributo se refere foi arbitrariamente escolhida marcando-se todas as células cujo valor de ordstrmax fosse igual ou superior a k. A Figura A.8 exemplifica o cálculo deste atributo considerando k = 3. A partir da grade LDD, identifica-se a célula à jusante (em verde) de uma determinada célula (em vermelho) que pertença à rede de drenagem. Calcula-se então a diferença altimétrica entre estas células. Por definição, todas as células pertencentes à rede de drenagem possuem valor de hand nulo.

↘	↘	↘	↘	↘	→	▪	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

74	99	94	59	20	14	0	12
76	68	68	54	19	0	6	35
50	32	28	31	16	0	25	58
48	13	0	10	0	14	48	77
60	24	14	0	3	29	67	97
80	64	41	11	0	53	50	101
112	80	50	8	4	42	78	106
114	78	35	26	10	50	84	95

hand_ordstrmax3

Figura A.8 – Exemplo de cálculo do atributo hand_ordstrmax3. As células marcadas em azul possuem valores de ordstrmax igual ou superior a 3. As setas em vermelho mostram o caminho que conecta a célula vermelha à verde.

d) Desnível ao k^{ésimo} vizinho à jusante - difz

Este atributo refere-se a diferença altimétrica entre uma célula e sua k^{ésima} vizinha à jusante seguindo a direção de fluxo. A Figura A.9 exemplifica o cálculo deste atributo considerando k = 4. A partir da grade LDD, identifica-se a célula à jusante (em verde) distante em 4 passos de uma determinada célula (em vermelho). Calcula-se então a diferença altimétrica entre estas células.

↘	↘	↘	↘	↘	→	▪	←
↘	↘	↘	↘	→	↗	↑	↖
↘	↘	↓	↘	↗	↑	↖	↖
→	→	↘	→	↗	↑	↖	↖
↗	↗	→	↗	↑	↖	↖	↖
↗	↗	↗	↑	↖	↖	↖	←
↗	↘	↗	↗	↑	↖	←	↖
→	→	↗	↗	↑	↖	↖	↖

LDD

200	218	208	173	134	127	113	125
202	194	187	168	133	114	119	148
176	158	154	150	130	115	139	171
174	139	126	129	119	129	163	191
186	150	135	121	122	148	182	212
206	185	162	132	125	172	169	220
233	205	171	133	129	167	203	225
239	203	160	151	135	175	209	220

MDE

79	103	95	60	21	14	0	12
83	75	73	55	20	1	6	35
57	43	39	37	17	2	26	58
55	24	12	16	6	16	50	78
67	35	21	8	9	35	69	99
87	70	48	18	11	58	55	105
114	84	56	18	14	52	84	110
114	82	41	32	16	56	90	99

difz4

Figura A.9 – Exemplo de cálculo do atributo difz. As setas em vermelho mostram o caminho que conecta a célula vermelha à verde.