

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

**MAPEAMENTO DIGITAL DE SOLOS COM USO DE ÁRVORES DE DECISÃO
NA MICROBACIA CÓRREGO TARUMÃZINHO, ÁGUAS FRIAS, SC**

José Janderson Ferreira Costa
(Dissertação)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO

**MAPEAMENTO DIGITAL DE SOLOS COM USO DE ÁRVORES DE DECISÃO
NA MICROBACIA CÓRREGO TARUMÃZINHO, ÁGUAS FRIAS, SC**

José Janderson Ferreira Costa
Engenheiro Agrônomo (UFRA)

Dissertação apresentada como
um dos requisitos à obtenção do
grau de Mestre em Ciência do Solo

Porto Alegre (RS), Brasil
Agosto de 2016

CIP - Catalogação na Publicação

Ferreira Costa, José Janderson

Mapeamento digital de solos com uso de árvores de decisão na microbacia Córrego Tarumãzinho, Águas Frias, SC / José Janderson Ferreira Costa. -- 2016. 70 f.

Orientador: Élvio Giasson.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS, 2016.

1. Pedologia. 2. Mapeamento digital de solos. 3. Predição de classes de solos. 4. Desagregação. I. Giasson, Élvio, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

JOSÉ JANDERSON FERREIRA COSTA

MAPEAMENTO DIGITAL DE SOLOS COM USO DE ÁRVORES DE DECISÃO
NA MICROBACIA CÓRREGO TARUMÃZINHO, ÁGUAS FRIAS, SC

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência do Solo da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul, como requisito parcial para obtenção do título de Mestre em Ciência do Solo.

Prof. Orientador: Dr. Élvio Giasson

Aprovada em 03 de Agosto de 2016
Homologada em 06 de Dezembro de 2016

BANCA EXAMINADORA:

Prof. Dr. Paulo César do Nascimento
UFRGS

Prof. Dr. Carlos Gustavo Tornquist
UFRGS

Prof. Dr. Egon Klamt

AGRADECIMENTOS

Primeiramente a Deus pelas oportunidades concedidas e por estar sempre presente em minha vida.

Aos meus pais, José Joaquim e Rosinete Ferreira, por todo o incentivo, conselho, carinho e por serem exemplos de pessoas honestas e trabalhadoras.

Ao meu orientador, professor Élvio Giasson, pela orientação, contribuição na dissertação e pela confiança na execução do estudo.

Ao programa de Pós-Graduação em Ciência do Solo (UFRGS), por meio dos seus professores e funcionários, pela oportunidade de crescimento profissional e pelo apoio técnico e científico.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pela concessão da bolsa de estudo.

Aos professores da banca Paulo César do Nascimento, Carlos Gustavo Tornquist e Egon Klamt pelas contribuições e aprimoramento do estudo.

À Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina, no nome de Elisângela Benedet, pela concessão de parte do material utilizado no estudo.

Aos professores, funcionários e amigos da E.M.E.F. Vilma de Nazaré Mendes que foi onde tudo começou e certamente tem uma enorme contribuição.

Aos professores, funcionários e amigos da Universidade Federal Rural da Amazônia pela oportunidade e crescimento profissional.

Aos meus colegas e amigos do grupo de pesquisa: Alcinei, Elisângela, Israel, Benito, Eliana, Joelma e Agi pela troca de ideias, aprendizado, parceria e pelo ótimo espírito de equipe.

A minha namorada Monique Nascimento, pelo amor, carinho, companhia e por proporcionar momentos de muita felicidade ao longo da Graduação e Pós-Graduação.

As minhas irmãs Janilce Ferreira e Jacielle Ferreira, pelo carinho, amor, ensinamentos e por compreenderem minha ausência.

Ao meu amigo Cristiano Zimmer pela recepção em Porto Alegre e pela amizade desde o período de Intercâmbio.

Muito obrigado!!!

MAPEAMENTO DIGITAL DE SOLOS COM USO DE ÁRVORES DE DECISÃO NA MICROBACIA CÓRREGO TARUMÃZINHO, ÁGUAS FRIAS, SC^{1/}

Autor: José Janderson Ferreira Costa
Orientador: Prof. Dr. Élvio Giasson

RESUMO

O conhecimento detalhado sobre a distribuição espacial dos solos é fundamental para o monitoramento ambiental e gerenciamento da produção agrícola. Novas técnicas de modelagem espacial estão sendo utilizadas para melhor aproveitar os dados de solos existentes como a desagregação espacial dos polígonos a partir do refinamento das unidades de mapeamento combinadas em que dois ou mais tipos de solos estão representados. Este trabalho foi dividido em dois estudos. O estudo 1 teve como objetivo realizar a comparação de três algoritmos de árvores de decisão (AD) para prever a ocorrência das unidades fisiográficas (UFs). Foram derivadas a partir do modelo digital de elevação (MDE) 12 variáveis ambientais utilizando o System for Automated Geoscientific Analyses (SAGA GIS). Os algoritmos testados foram: *J48*, *Simple Chart*, *BFTree*. As ADs foram construídas no programa Weka 3.6.3. As acurácias obtidas usando os três algoritmos foram semelhantes, sendo que o *J48* foi superior, mostrando-se útil para prever a ocorrência de unidades fisiográficas. O estudo 2 teve como objetivo desagregar os polígonos do mapa fisiográfico para individualizar as classes de solos até o segundo nível categórico. Para realizar a desagregação do mapa fisiográfico foram utilizadas 3 variáveis ambientais preditoras, derivadas do MDE, sendo que o mapa de geoforma foi derivado utilizando um conjunto de programas chamado LandMapr toolkit®. Os limites das UFs, os mapas de elevação, declividade e geoforma foram submetidos a uma tabulação cruzada para identificar em cada UF as combinações existentes entre suas classes. O uso da desagregação possibilitou a individualização e a espacialização das classes de solos, sendo útil para a produção de mapas de solos mais detalhados.

1/ Dissertação de Mestrado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (70 p) Agosto, 2016. Trabalho realizado com apoio financeiro do CNPq.

DIGITAL SOIL MAPPING USING DECISION TREES IN TARUMÃZINHO STREAM WATERSHED, ÁGUAS FRIAS, SC^{2/}

Author: José Janderson Ferreira Costa

Adviser: Prof. Dr. Élvio Giasson

ABSTRACT

Detailed knowledge about the spatial distribution of the soil is essential for environmental monitoring and management of agricultural production. New spatial modeling techniques have used to make better use of data from existing soils and the polygons spatial disaggregation from the refinement of combined mapping units. This work was is divided two studies. The study 1 aimed to carry out a comparison of three decision trees algorithms to predict the occurrence of physiographic units (UFs). The System for Automated Geoscientific Analyses (SAGA GIS) was used to derive 12 environmental covariates from digital elevation model. Three algorithms were tested: J48, Simple Chart, BFTree. The program Weka 3.6.3 was used to built the decision trees. The accuracies obtained using the three algorithms were similar. The J48 accuracy was higher, proving to be useful to predict the occurrence of physiographic units. The study 2 aimed to disaggregate the physiographic map polygons, to individualize the soil classes by the second categorical level. Three environmental covariates were used to carry out the polygon disaggregation of physiographic unit map. The environmental covariates was derived from the MDE, and the map of geoforma was elaborated by a set of programs called LandMapr toolkit©. The limits of the UFs, maps of elevation, slope and geoforma underwent a cross-tab to identify existing combinations between classes, by each UF. The polygon disaggregation made possible the individualization and the spatialization of soils classes, being useful to production of more detailed soil maps.

2/ Master dissertation in Soil Science. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (70 p.) August, 2016. Work carried out with financial support from CNPq.

SUMÁRIO

	Página
1. INTRODUÇÃO GERAL.....	1
2. REVISÃO BIBLIOGRÁFICA.....	4
2.1. Relação solo-relevo	4
2.2. Levantamentos de solos	6
2.3. Mapeamento digital de solos	8
2.4. Algoritmos de árvores de decisão.....	11
2.5. Variáveis ambientais preditoras.....	13
2.5.1. Principais variáveis preditoras.....	14
2.6. Desagregação de polígonos de mapa	16
3. ESTUDO 1 – PREDIÇÃO DA OCORRÊNCIA DE UNIDADES FISIAGRÁFICAS COM USO DE TRÊS ALGORITMOS DE ÁRVORES DE DECISÃO	19
3.1. Introdução.....	19
3.2. Materiais e Métodos.....	21
3.2.1. Área de estudo.....	21
3.2.2. Identificação das UFs e solos dominantes.	22
3.2.3. Metodologia.....	25
3.3. Resultados e discussão	28
3.3.1. Algoritmos de Árvores de Decisão.....	34
3.4. Conclusões	43
4. ESTUDO 2 – INDIVIDUALIZAÇÃO DE CLASSES DE SOLOS POR DESAGREGAÇÃO DE POLÍGONOS DE MAPA FISIAGRÁFICO.	44
4.1. Introdução.....	44
4.2. Materiais e métodos.....	45
4.2.1. Área de estudo.....	45
4.2.2. Metodologia.....	47
4.3. Resultados e discussão	51
4.4. Conclusões	60
5. CONSIDERAÇÕES FINAIS	60
6. REFERÊNCIAS.....	62

RELAÇÃO DAS TABELAS

Tabela 1. Descrição das unidades fisiográficas e os solos ocorrentes na microbacia Córrego Tarumãzinho. Fonte: Epagri (2004).	25
Tabela 2. Matriz de erro do Algoritmo J48, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.	35
Tabela 3. Matriz de erro do Algoritmo SimpleChart, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.	37
Tabela 4. Matriz de erro do Algoritmo BFTree, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.	39
Tabela 5. Comparação dos resultados de acurácia, índice Kappa, tamanho da árvore de decisão e unidades fisiográficas preditas pelos três algoritmos.	42
Tabela 6. Descrição das unidades fisiográficas e os solos ocorrentes na microbacia Córrego Tarumãzinho. Fonte: Epagri (2004).	49
Tabela 7. Regras de individualização das classes de solos nos polígonos de unidades fisiográficas combinadas, a partir de mapas de classes de declividades, elevação, geofoma e os limites dos polígonos do mapa fisiográfico.	52
Tabela 8. Valores de erro absoluto e erro relativo entre as áreas do mapa fisiográfico e a soma das áreas de classes de solos ocorrentes nas UFs, representadas no mapa de desagregação.	54
Tabela 9. Área ocupada pelo mapa fisiográfico em relação a área ocupada pelas classes de solos desagregadas ocorrentes nas unidades fisiográficas.	58
Tabela 10. Medidas de acurácia da concordância dos perfis georreferenciados com a localização de pixels individuais da microbacia Córrego Tarumãzinho.	59

RELAÇÃO DAS FIGURAS

Figura 1. Localização da microbacia Córrego Tarumãzinho no oeste do Estado de Santa Catarina. Fonte: Epagri (2004).	21
Figura 2. Visão 3D no Google Earth das unidades fisiográficas da microbacia Córrego Tarumãzinho.	23
Figura 3. Mapa fisiográfico da microbacia Córrego Tarumãzinho. Fonte: Epagri (2004). Componentes: Cume Erosional (CE), Encosta Coluvial Erosional (ECE), Encosta Erosional (EE), Encosta Erosional Coluvial (EEC), Encosta em Patamar (EP), Fundo de Vale Coluvial Aluvial (FVCA).....	24
Figura 4. Distância diagonal da drenagem da microbacia Córrego Tarumãzinho.	29
Figura 5. Área de contribuição da microbacia Córrego Tarumãzinho.	31
Figura 6. Altitude da microbacia Córrego Tarumãzinho.	32
Figura 7. Declividade do terreno na microbacia Córrego Tarumãzinho. ...	33
Figura 8. Mapa fisiográfico predito produzido pelo algoritmo J48 e mapa convencional da microbacia Córrego Tarumãzinho.	34
Figura 9. Mapa fisiográfico predito produzido pelo algoritmo Simple Chart e mapa convencional da microbacia Córrego Tarumãzinho.	38
Figura 10. Mapa fisiográfico predito produzido pelo algoritmo BFTree e mapa convencional da microbacia Córrego Tarumãzinho.	41
Figura 11. Localização da microbacia Córrego Tarumãzinho no oeste do Estado de Santa Catarina. Fonte: Epagri (2004).	46
Figura 12. Mapa fisiográfico da microbacia Córrego Tarumãzinho. Fonte: Epagri (2004). Componentes: Cume Erosional (CE), Encosta Coluvial Erosional (ECE), Encosta Erosional (EE), Encosta Erosional Coluvial (EEC), Encosta em Patamar (EP), Fundo de Vale Coluvial Aluvial (FVCA).....	48
Figura 13. Mapa de unidades taxonômicas. (a) Detalhe de uma porção das UFs combinadas no mapa fisiográfico; (b) Detalhe de uma porção do mapa fisiográfico desagregado.	53
Figura 14. Distribuição da área em relação as classes de solos representadas em unidades de mapeamento simples no mapa desagregado da microbacia Córrego Tarumãzinho.	56
Figura 15. Relação entre a área do mapa fisiográfico e a área do mapa desagregado com as classes de solos ocorrentes representas de forma individual.	57

RELAÇÃO DAS ABREVIATURAS

- AD:** Árvore de decisão.
- AG:** Acurácia geral.
- AM:** Acurácia do mapeador.
- AMM:** Área mínima mapeável.
- CLI:** *Canada Land Inventory*.
- CSV:** *Comma Separated Values*.
- EMBRAPA:** Empresa Brasileira de Pesquisa Agropecuária.
- FAO:** *Food and Agriculture Organization*.
- GPS:** *Global Positioning System*.
- IBGE:** Instituto Brasileiro de Geografia e Estatística.
- IUT:** Índice de umidade topográfica.
- MDE:** Modelo Digital de Elevação.
- MF:** Mapa Fisiográfico.
- MDS:** Mapeamento Digital de Solos.
- MDT:** Modelo Digital do Terreno.
- MNT:** Modelo Numérico do Terreno.
- RMS:** Erro médio quadrado.
- SAD69:** *South American Datum 1969*.
- SC:** Santa Catarina.
- SiBCS:** Sistema Brasileiro de Classificação de Solos.
- SIG:** Sistema de Informação Geográfica.
- SIRGAS2000:** Sistema de Referência Geocêntrico para as Américas.
- SRTM:** *Shuttle Radar Topography Mission*.
- UF:** Unidade Fisiográfica.
- UM:** Unidade de Mapeamento.
- UT:** Unidade Taxonômica.
- WEKA:** *Waikato Environment for Knowledge Analysis*.

1. INTRODUÇÃO GERAL

Alguns sistemas de uso da terra vêm provocando um acelerado processo de degradação dos solos, ocasionando a alteração dos atributos químicos, físicos e biológicos. Dada a importância desse recurso, é necessário utilizá-lo de forma sustentável para o adequado funcionamento dos ecossistemas, produzindo alimentos e outros recursos sem causar desequilíbrio à vida do planeta. Com o crescimento populacional e a maior demanda pela produção de alimentos, fibras e biocombustíveis, muitos órgãos nacionais e internacionais têm demonstrado preocupações pela forma como o solo está sendo utilizado e promovem iniciativas de conscientização e compreensão da importância dos solos para a segurança alimentar e funções essenciais dos ecossistemas (FAO; ITPS, 2015).

Apesar do crescente interesse da sociedade, a conservação do solo só poderá ser alcançada com o conhecimento atualizado e detalhado da distribuição espacial dos solos na paisagem e suas propriedades. Uma importante fonte de informações são os levantamentos dos solos, utilizados para o planejamento do uso, manejo e conservação do solo. No Brasil, ao longo das últimas décadas houve redução dos investimentos destinados aos levantamentos de solos por parte dos governos e das agências financiadoras. Essas informações se tornaram escassas pela restrição de financiamentos para a ciência básica e, adicionalmente, muitas vezes os dados de solos gerados encontram-se em escala pequena, restringindo sua utilidade (Sanchez et al., 2009; Chagas et al., 2010).

Frente a esse desafio, novos métodos e ferramentas para o levantamento e mapeamento de solos têm sido propostos na última década (McBratney et al., 2003; Carvalho et al., 2009; Giasson et al., 2011; ten Caten et al., 2011a), permitindo o conhecimento da distribuição espacial de classes e/ou propriedades dos solos a partir de variáveis ambientais e dados de solos existentes, disponibilizando aos diferentes usuários informações que facilitam a tomada de decisão quanto ao manejo e uso do solo.

O Sistema de Informação Geográfica (SIG) é uma ferramenta que permite analisar e interpretar dados relativos à superfície terrestre, sendo

integrado por um conjunto de *software* e *hardware* desenvolvidos, basicamente, para a aquisição, processamento e uso de dados cartográficos. É no SIG que se realiza o mapeamento digital de solos (MDS), uma técnica que permite estabelecer relações matemáticas entre variáveis ambientais preditoras e as classes de solos, gerando informações contínuas sobre a distribuição espacial dos solos (Coelho & Giasson, 2010; McBratney et al., 2003).

A utilização do MDS possibilita desenvolver e disponibilizar as informações espaciais através da análise das relações entre solo-paisagem, permitindo identificar os padrões de ocorrência de diferentes tipos de solos usando como base os fatores de formação do solo, tornando-se uma técnica econômica e rápida para geração de informações dos solos a partir de modelos numéricos do terreno ou da superfície (McBratney et al., 2003).

O MDS é baseado em modelos de predição que utilizam variáveis ambientais para treinamento de algoritmos, modelagem e extrapolação da ocorrência de solos para aplicação em áreas contíguas, com características ambientais semelhantes. Vários métodos estatísticos são utilizados para gerar os modelos ou regras de classificação que posteriormente serão aplicadas para produção dos mapas digitais de solos. Dentre esses métodos podemos destacar os modelos de regressões logísticas múltiplas multinomiais (Giasson et al., 2006), modelos de redes neurais artificiais (Sirtoli, 2008), modelos logísticos com aplicação de componentes principais (ten Caten et al., 2009) e o modelo de árvores de decisão (Giasson et al., 2013).

Estudos tem evidenciado que o uso de árvores de decisão (AD) apresenta bons resultados no MDS e tem a vantagem de ser um modelo não paramétrico, quantitativo e que utiliza variáveis quantitativas ou categóricas (Basgalupp, 2010). Outra vantagem do modelo de AD é a capacidade para discriminação de classes de solos (Coelho & Giasson, 2010; Giasson et al., 2011).

Outra técnica que visa gerar mais informações sobre solos é a desagregação poligonal de mapas convencionais de solos. Essa metodologia foi utilizada por Häring et al. (2012), Kerry et al. (2012), Odgers et al. (2014a) e Subburayalu et al. (2014) e tem como finalidade desagregar unidades de mapeamento (UM) de mapas com menor detalhamento, através do uso de geotecnologias, gerando maior número de UM de solos mais homogêneas. Isso

tem sido possível devido ao uso de GIS e a modelagem de paisagem a partir de Modelo Numérico do Terreno (MNT). O uso dessas tecnologias e a experiência adquirida pelos pedólogos, se bem retratada nos relatórios de levantamentos de solos, criam essa possibilidade de maior detalhamento nos mapas de solos sem trabalhos adicionais de campo.

Dentro desse contexto da necessidade demonstrada de mais informações de solos e com as possibilidades elencadas pelas novas tecnologias empregadas na modelagem quantitativa e no MDS, pode-se esperar a geração de mais informações sobre solos. Todavia, embora as tecnologias utilizadas já sejam consagradas, vários aspectos metodológicos têm sido recentemente desenvolvidos e muitos outros ainda necessitam de pesquisas que possibilitem o desenvolvimento metodológico adequado.

2. REVISÃO BIBLIOGRÁFICA

2.1. Relação solo-relevo

Para os pesquisadores do solo, o interesse em conhecer e entender as relações entre o solo e o relevo é uma constante e teve início com as primeiras abordagens feitas por Dokutchaev, no século XIX (1883), reforçadas e amplamente divulgadas no início do século XX (1941) por Jenny e outros autores (Johnson & Schaetzl, 2015). Nessa abordagem fatorial, Dokutchaev entende o solo como uma função de quatro fatores (clima, organismos, relevo e material de origem) que interagem através do tempo. Ele priorizou o clima como um fator importante de formação dos solos e reconheceu também a importância do subsolo, da vegetação e da fauna.

Como um dos fatores de formação do solo, o relevo pode ser definido como o conjunto de desníveis da crosta originados pelas forças exógenas, tais como o clima e a vegetação, bem como, as forças endógenas, a tectônica e a geologia, expressados em várias escalas (Muñoz, 2009). As diferentes formas do relevo, a ocorrência e duração dos processos pedológicos e as características do material de origem determinam, muitas vezes, o tipo e distribuição dos solos na paisagem (Wysocki et al., 2012). Solos encontrados em posições mais altas na paisagem (por exemplo interflúvios), apresentam maior desenvolvimento e melhor drenagem em comparação a solos desenvolvidos em sopé, fator diretamente relacionado com a posição do solo na paisagem (Campos et al., 2011).

No estudo do terreno como um sistema, a principal aplicação do relevo ao mapeamento das variáveis do terreno é a abordagem quantitativa e a divisão do terreno em unidades de características qualitativamente homogêneas (Briggs, 1985; Muñoz, 2009). O sistema terreno engloba o solo, a vegetação e as geoformas, sendo que esta última é na prática a base preponderante da classificação. Isto porque as geoformas resultam da ação da maioria dos fatores que controlam a gênese e distribuição do solo (Ollier, 1977).

Em relação aos atributos pedológicos, o relevo é um importante condicionador das propriedades do solo (McBratney et al., 1999), por interferir no escoamento superficial, nos movimentos verticais e horizontais da água e,

consequentemente, nos processos pedogenéticos. Esses processos modificam a disponibilidade de nutrientes em escalas locais e promovem variações na mineralogia do solo (Bui, 2004).

Como um fator pedogenético dominante em nível local, o relevo gera certas modificações nas propriedades dos solos que variam ao longo das vertentes, como a granulometria, as bases trocáveis e outras características que determinam o desenvolvimento do perfil do solo, sendo que essa variação é interpretada como sendo o resultado da distribuição e acumulação de água (Jenny, 1941; MAcMillan et al., 2008).

A pedologia sempre considerou o relevo um fator importante para explicar os processos de formação dos solos, uma vez que o desenvolvimento de diferentes tipos de solos em diferentes posições no relevo indica que os processos de formação também são diferentes para cada segmento de uma vertente (Gerrard, 1981). Esse contraste no desenvolvimento dos solos está relacionado a vários fatores condicionados pelo relevo, por exemplo, a variação da drenagem, dado que as áreas mais altas são bem drenadas em relação as partes mais baixas, sendo que, essa diferença de drenagem é uma das responsáveis pelas mudanças graduais nas características e no desenvolvimento dos perfis de solos (Embrapa, 2013).

Para melhorar a compreensão da atuação do relevo na formação dos solos, foi proposto o modelo hipotético dos processos geomorfológicos dominantes em uma vertente apresentado por Christofolletti (1980). O autor dividiu o modelo em nove unidades de superfície: 1) Seção plana (interflúvio), onde ocorrem os processos pedogenéticos associados com o movimento vertical subsuperficial da água; 2) Declive de infiltração, ocorre o processo de eluviação mecânica e química pelo movimento lateral subsuperficial da água; 3) Declive convexo, ocorre a reptação (rastejamento) e a formação de terracetes; 4) Escarpa ou frente de queda, ocorre os deslizamentos, desmoronamentos e o desgaste químico e físico do solo; 5) Declive intermediário ou meia encosta, ocorre o transporte de material pelos movimentos coletivos do solo, formação de terracetes, ação da água superficial e subsuperficial; 6) Sopé coluvial, ocorre a deposição de material pelos movimentos de solo e escoamento superficial; formação de leques; transporte de material, ação subsuperficial da água; 7) Declive aluvial, ocorre a deposição aluvial e processos oriundos do movimento

subsuperficial da água; 8) Margem do curso d'água, ocorre os processos de erosão e perda dos solos; 9) Leito do curso d'água, ocorre o transporte de material para jusante pela ação da água superficial, gradação e erosão periódicas.

Dada a importância do relevo na distribuição espacial das classes de solos, é justificável seu emprego nos levantamentos de solos e nos estudos com MDS aplicados na predição de ocorrência de classes de solos.

2.2. Levantamentos de solos

Um levantamento de solos é um inventário cartográfico dos solos de uma determinada área, é composto por um mapa com a distribuição espacial dos solos e um relatório técnico contendo a classificação, a descrição morfológica e as informações de análises químicas e físicas dos perfis representativos dos solos (Bui, 2004). As informações contidas em um levantamento auxiliam no desenvolvimento do plano de gestão das terras, servem como bases de estudo da viabilidade técnica e econômica das atividades desenvolvidas no meio ambiente e são importantes na tomada de decisão e gerenciamento da produção agrícola e de outras atividades.

Os levantamentos de solos existentes no Brasil cobrem grandes áreas e foram realizados de forma convencional, sendo executados com baixo nível de detalhamento, em nível exploratório ou de reconhecimento (1:5.000.000 a 1:100.000), resultando em mapas de solos em escala pequena, carecendo de informações com maior detalhe em nível local para o planejamento do uso e manejo de solos (Mendonça-Santos & Santos, 2007).

Os levantamentos exploratórios (escala 1:750.000 ou menor) cobrem 94% do território nacional, no entanto, esse nível de detalhamento não permite o planejamento agrícola e conservacionista. Já os levantamentos de reconhecimento de alta intensidade (escala 1:60.000 a 1:100.000), os semidetalhados (escala 1:50.000 ou maior) e os detalhados (escala maior ou igual a 1:20.000) abrangem apenas 1,71% do território nacional, ficando restrito a pequenas áreas (Santos et al., 2013).

Em muitas áreas, onde as mudanças no uso e cobertura do solo são motivadas principalmente pelas atividades agropecuárias, os dados de solos utilizados para atender os objetivos de planejamento e ocupação de forma

ordenada são insuficientes e generalizados (escala 1:250.000 ou menor), ocasionando problemas de impacto ambiental e degradação dos solos, além de baixa produtividade e perda de solo pela erosão (Manzatto, 2002).

Extensas áreas no território não têm informações completas dos solos em escalas que satisfaçam as necessidades para tentar solucionar parte dos problemas atuais de utilização do solo (Mendonça-Santos & Santos, 2007). Levantamentos de solos em escala mais detalhada são necessários para suprir a demanda por informações adequadas ao planejamento e gestão de projetos agrícolas e ambientais (Menezes, 2013).

A carência de dados com maior detalhe dos solos é relatada em diversos trabalhos ao longo dos anos. Segundo Santos (1992), na década de 1960, a produção de mapas de solos das regiões norte e centro-oeste marcou o início de um novo período de treinamento de campo dos cientistas do solo no Brasil e nos Estados Unidos, contribuindo para um aumento substancial de profissionais para realizarem os levantamentos. No entanto, esses levantamentos foram publicados em diferentes níveis e escalas, mas predominantemente nos níveis exploratórios e de reconhecimento, tornando-se insuficientes para trabalhos que exigiam o conhecimento do solo em escala maior (Mendonça-Santos & Santos, 2007).

Nos primeiros levantamentos de solos no Brasil na década de 1960, optou-se por realizar levantamentos em escala pequena, basicamente, pela escassez de recursos financeiros, carência de pessoal qualificado e grande extensão da área a ser mapeada, dentre outros fatores. O processo de reorganização e incorporação do extinto Serviço Nacional de Levantamento e Conservação de Solos (atual Centro Nacional de Pesquisa de Solos) responsável pelos levantamentos, também inviabilizaram a execução e atualização dos levantamentos em escala maior (Ibañez et al., 1993; Chagas, 2006).

Décadas depois, os institutos continuaram realizando levantamentos a nível de reconhecimento (Embrapa, 1981) com a finalidade de identificar áreas apropriadas à instalação de projetos de colonização, porém, insuficientes em termos de escala para atender a demanda e expectativas recentes, necessitando de atualização de técnicas para produzir mapas mais completos.

Apesar da tentativa, pouco se avançou na produção de mapas com informações mais detalhadas dos solos. Além dos fatores elencados acima, o que torna essas informações escassas são, principalmente, a redução do apoio institucional para pedologia pelas agências financiadoras, além das mudanças ocorridas nas próprias instituições de pesquisa de solos, resultando na restrição do nível de detalhamento e a expansão dos levantamentos, deixando muitas regiões mapeadas com baixa resolução e uma escassez generalizada de mapas detalhados. Na década de 1980, poucos trabalhos foram realizados, período no qual os levantamentos de solos praticamente estagnaram (McBratney et al., 2003; Hartemink & McBratney, 2008).

Segundo Chagas et al. (2010), até o presente momento poucas áreas do território brasileiro apresentam levantamentos de solos em média escala (entre 1:100.000 e 1:50.000) e com detalhe adequado para a realização de planejamento. Sem o conhecimento suficiente sobre os solos, as empresas agrícolas (principalmente dos setores canavieiro e florestal) estão realizando seus próprios levantamentos nas áreas de produção, com intuito de identificar o real potencial produtivo das terras, além de garantir o crescimento agrícola de forma sustentável.

Nesse contexto, a busca por mapas de solos que atendam as expectativas recentes, é um desafio para os cientistas (Menezes, 2013), sendo necessário o uso de novas técnicas ao mapeamento de solos que aproveitem os dados históricos de levantamentos convencionais. Logo, a inserção de técnicas computacionais nos estudos ambientais torna o MDS uma técnica que pode estimular uma nova fase para produzir mapas de solos em escala adequada e ajustando os procedimentos através da utilização de novas ferramentas (McBratney et al., 2012; Omuto, et al., 2013).

2.3. Mapeamento digital de solos

O mapeamento digital de solos (MDS) surgiu como alternativa para a criação de mapas com maior resolução espacial e visa complementar os levantamentos de solos, no intuito de suprir a demanda de informações de solos em áreas não mapeadas. O MDS é sustentado pela tecnologia da informação e dados de campo para a inferência das variações espaciais dos tipos de solos e suas propriedades, utiliza dados históricos de solos e variáveis ambientais

correlacionadas, como a declividade e a curvatura, dentre outras (Lagacherie & McBratney, 2007).

Os primeiros trabalhos com MDS iniciaram em meados de 1970, mas foi na década de 1980 que passou a ser usado com maior frequência devido aos avanços tecnológicos nas áreas de sensoriamento remoto, da computação, da estatística e modelagem, dos sistemas de informações geográficas, dos sistemas de posicionamento global, dos sistemas de medição (tais como espectroscopia de infravermelho), e mais recentemente, através do acesso a banco de dados disponíveis na internet (Minasny & McBratney, 2016).

Embora o MDS tenha como base o conhecimento acumulado e o uso de tecnologias disponíveis, a técnica permanece fundamentada no primeiro modelo preditivo proposto por Jenny (1941), na qual formulou a equação de formação do solo. No que se refere à abordagem clássica, ou método de correlação ambiental, a inferência é embasada nos fatores de formação do solo: clima (cl), organismos (o), relevo (r), material parental (p) e tempo (t), sendo o tempo uma variável independente, enquanto, as demais são dependentes e correlacionadas.

Uma nova abordagem da equação de Jenny foi proposta por McBratney et al. (2003), inserindo dois novos fatores no conjunto de variáveis preditoras: a posição espacial e o próprio solo. O primeiro fator foi incluído porque os autores consideram que o solo pode ser predito por outros atributos do solo na mesma posição espacial ou posições vizinhas. O segundo fator foi proposto porque o solo pode ser predito a partir de seus próprios atributos através de equações de pedotransferência. A nova equação derivada do modelo de Jenny (1941) foi denominada como modelo *scorpan*, onde solo (s), clima (c), organismos (o), relevo (r), material parental (p) tempo (a) e sua posição no espaço (n). No modelo *scorpan*, a localização espacial é vista como um atributo que confere poder preditivo nos modelos, considerando o vasto conjunto de geotecnologias disponíveis.

Assim como nos levantamentos convencionais, grande parte dos trabalhos com MDS baseiam-se nas relações entre os solos e as características da paisagem. Segundo Sanchez et al. (2009), as etapas gerais do MDS são: (1) reunião e junção de dados relativos a clima, geologia, relevo, além de informações disponíveis em relatórios e mapas de solos sobre sua distribuição

espacial; (2) mapeamento de classes e propriedades do solo através da aplicação de relações matemáticas entre essas e os fatores de formação do solo; (3) aplicação das informações mapeadas com vistas a gerar novas informações, as quais são mais difíceis de serem medidas do que estimadas; (4) a informação gerada na etapa anterior é confrontada com as atividades antrópicas, gerando mapas de conflitos, riscos e possibilidades e, por fim, (5) a geração de um conjunto de medidas e recomendações baseado no conhecimento da distribuição espacial do solo.

Como exemplo de aplicação do MDS no Brasil, Teske et al. (2014) compararam modelos de predição de classes de solos gerados a partir de dados de solos e atributos do terreno derivados de diferentes modelos digitais de elevação (MDEs). Os resultados indicaram que os modelos preditores mais acurados e com maior número de UMs estimadas foram os gerados a partir dos MDEs com resolução espacial de 90 m (SRTM v4.1 e CN90).

Crivelenti et al. (2009) também utilizaram dados e informações de solos contidas em um mapa pedológico na escala 1:100.000 e técnicas de mineração de dados para a predição de classes de solos. A partir do MDE, os autores derivaram as variáveis ambientais declividade, curvatura planar, curvatura do perfil, área de contribuição e distância diagonal de drenagem, permitindo a elaboração de mapas digitais semelhantes aos mapas de solos convencionais.

Coelho & Giasson (2010) testaram e compararam métodos de classificação em estágio único (regressões logísticas múltiplas multinomiais e Bayes) e algoritmos de árvores de decisão (*CART*, *J48* e *LMT*) para produção de mapas pedológicos com legenda original e simplificada. As variáveis do terreno e o mapa convencional de solos foram usadas para o treinamento dos algoritmos. O modelo de AD teve maior poder de discriminação em relação aos métodos de classificação em estágio único e também apresentou melhor resultado para mapeamento supervisionado de solos com variáveis geomorfométricas.

Giasson et al. (2013) avaliaram cinco algoritmos de AD na Bacia do Lageado Grande, em uma área de 532 km². Os algoritmos *Simple Chart*, *J48* e *BFTree* geraram AD com capacidade ligeiramente superior na produção de

mapas preditores semelhantes ao original. O algoritmo que apresentou maior valor de acurácia geral média 55,3% foi o Simple Chart.

Apesar do crescente uso em trabalhos científicos, o modelo matemático de AD precisa ser melhor estudado, já que existem diversos algoritmos para geração das regras de classificação, e cada algoritmo pode apresentar resultados diferentes em relação a sua capacidade de predição das classes de solos, sendo necessário maiores estudos dos principais algoritmos na predição de unidades fisiográficas em escalas maiores.

2.4. Algoritmos de árvores de decisão

Boa parte das técnicas de mineração de dados tem como base a aprendizagem indutiva, onde um modelo é construído utilizando-se parte dos dados para o treinamento. A finalidade da abordagem indutiva é que o modelo treinado possa ser aplicado em áreas não amostradas (Rokach & Maimon, 2008).

O modelo de árvores de decisão (AD) tem sido frequentemente empregado nos estudos com MDS para gerar regras de classificação a partir de dados de solos e variáveis ambientais derivadas do MDE. Há várias razões por trás de sua visibilidade e ampla aplicabilidade, dentre elas, que a precisão de AD é comparável ou superior à precisão com outros modelos de classificação. Além disso, o modelo de AD gera um menor número de equações em comparação a outros modelos, o que facilita a implementação no SIG, minimizando as chances de erro (Maia, 2013).

O modelo de AD pode utilizar diversos algoritmos, tais como *J48* (Quinlan, 1993), *SimpleChart* (Breiman et al., 1984) e *BFTree* (Shi, 2007), dentre outros. Na construção da AD, o algoritmo *J48* avalia a existência ou significância dos atributos que melhor subdividem o conjunto das amostras em subconjuntos homogêneos. A homogeneidade pode ser definida por uma série de medidas, as quais constituem a principal diferença entre os diversos algoritmos (Bui & Moran, 2003; Scull et al., 2005). No *J48* a construção da AD se dá do topo para a base, simplificando as informações em cada nó e o nó terminal constitui uma folha, um conjunto homogêneo segundo as regras de decisão. Utiliza-se como critério o ganho de informação obtida na escolha do atributo para subdivisão (Hall et al., 2009).

O *J48* produz um esquema de classificação, através de várias interações, onde os nós representam um teste em determinados atributos de entrada e as folhas representam a classificação, resultado dos casos encaminhados ao longo dos ramos correspondentes. A cada interação o algoritmo escolhe o atributo que melhor diferencia os dados com base no ganho de informação e separa o conjunto em duas partes até que restem apenas nós folhas, ou seja, respostas de classificação (Natale et al., 2015).

O algoritmo *SimpleChart* é uma implementação do algoritmo CART proposto por Breiman et al. (1984), e é caracterizado pela construção de ADs binárias, ou seja, cada nó interno tem exatamente duas árvores de saída. Uma característica importante do *SimpleChart* é a sua capacidade para gerar árvores de regressão. É um algoritmo não paramétrico desenvolvido para análise de dados mediante classificação e regressão. Sua vantagem é a capacidade de realizar pesquisa de relação entre os dados, prevendo o tratamento de variáveis dependentes discretas. O resultado deste algoritmo é sempre uma AD binária univariada que pode ser percorrida da sua raiz até às folhas, respondendo apenas a questões simples do tipo sim/não.

O algoritmo *BFTree* segue a mesma lógica de AD do algoritmo *J48*, a diferença é que o *BFTree* realiza uma busca no espaço de subconjuntos de atributos através de métodos que utilizam um número maior de dados, tornando-se uma desvantagem para esse algoritmo (Witten e Frank, 2005). Em cada nó interno da árvore é utilizado somente uma variável como teste, e a seleção da variável pode ser feita tanto pelo índice *Gini* como pelo critério de ganho de informação. O objetivo da utilização do *BFTree* é a construção de um modelo que utiliza os critérios de divisão para reduzir ao máximo a impureza (Wu et al. 2015). Adota-se um critério diferente para problemas de duas ou múltiplas classes quando a variável é qualitativa. Para problemas de duas classes, o procedimento é o mesmo adotado para o algoritmo CART, onde o método reduz o espaço de busca. Nos problemas de múltiplas classes, é utilizado um método heurístico que tem a finalidade de reduzir a impureza e a velocidade de busca (Tan et al., 2005).

Como observado, existem diferenças entre os diversos algoritmos de AD que necessitam ser melhor avaliados quanto ao desempenho. Sendo assim, os algoritmos de AD *J48*, *SimpleChart* e *BFTree* serão testados e avaliados

quanto a predição da ocorrência de unidades fisiográficas. Dessa forma espera-se que os resultados contribuam para o MDS no desenvolvimento de técnicas que auxiliem no mapeamento digital de unidades fisiográficas.

2.5. Variáveis ambientais preditoras

As variáveis ambientais preditoras derivadas do MDE são compreendidas como atributos topográficos (Wilson e Gallant, 2000; Chagas, 2006), do terreno (Giasson et al., 2006; ten Caten et al., 2009) ou do relevo (Ippoliti et al., 2005; Sirtoli et al., 2008). Também são descritas na literatura como variáveis geomorfométricas (Bispo et al., 2009) ou morfométricas (Hermuche et al., 2002).

De acordo com Pissara et al. (2004), as variáveis ambientais preditoras são propriedades do terreno que refletem o padrão de drenagem e do relevo, como o escoamento superficial da água e infiltração. Elas também expressam as mudanças na paisagem que estão associadas com a litologia, estrutura geológica e formação superficial dos elementos que compõe a superfície terrestre.

As variáveis ambientais mais empregadas no estudo com MDS estão relacionadas aos fatores de formação do solo, material de origem e, principalmente, ao relevo. Segundo ten Caten et al. (2012), a predominância das variáveis derivadas do relevo deve-se à forte influência do mesmo na formação dos solos e, também pela facilidade de obtenção dos dados.

As variáveis ambientais preditoras podem ser divididas em variáveis primárias e secundárias (Moore et. al., 1993). As variáveis primárias são derivadas diretamente do MDE e incluem variáveis como elevação, curvatura do perfil, curvatura planar, declividade, orientação de vertentes, comprimento do fluxo e área de contribuição (Sirtoli, 2008). Outra forma de obtenção das variáveis é a partir da combinação das variáveis primárias, que estão relacionadas com o processo de transporte de sedimentos e o movimento de água superficial e subsuperficial no solo, tais como o índice de umidade topográfica (IUT) e o índice de posição topográfica (TPI), entre outros (Wilson e Gallant, 2000; McBratney et al., 2003).

Entre as variáveis primárias, as que tem apresentado os melhores resultados para explicar a distribuição espacial dos solos na paisagem são a

elevação, a declividade, o comprimento de fluxo e a distância horizontal das redes hidrográficas, as quais têm se tornado objeto de estudo para o MDS (Behrens et al., 2010; ten Caten et al., 2012; Giasson et al., 2013).

2.5.1. Principais variáveis preditoras

A maioria das variáveis preditoras correspondem a medidas morfométricas do relevo, mas elas também descrevem direta ou indiretamente outros fatores de formação dos solos e, portanto, são úteis na predição de ocorrência do mapa fisiográfico. Sendo assim, todas as variáveis derivadas a partir do MDE serão utilizadas na predição da ocorrência das unidades fisiográficas.

A elevação é importante pois está relacionada aos gradientes climáticos locais (Wilson e Gallant, 2000) e tem influência direta com a temperatura, a qual é muito importante para o acúmulo de matéria orgânica. Teske et al. (2015), testaram e compararam os resultados das predições de ocorrência de classes de solos com dados oriundos de três esquemas de amostragem, concluíram que a elevação foi a principal variável preditora de ocorrência de solos, tanto no esquema de amostragem aleatória, como no esquema de amostragem estratificada.

A variação de elevação causa grande diferença de temperatura em pequenas distâncias, essa relação é especialmente importante para as regiões tropicais e subtropicais, onde a mudança de elevação provoca sensíveis alterações no clima, solo e vegetação natural (Fritzsons, 2008). Silva (2012) testou sete variáveis ambientais e concluiu que a elevação foi a variável ambiental com maior poder de predição de classes de solos.

A declividade representa a primeira derivada da altitude e corresponde a um plano tangente à superfície, ou seja, é a inclinação da superfície do terreno em relação à horizontal (Gallant e Wilson, 2000). Quanto maior a declividade, maior será o movimento de água na superfície do solo, influenciando o modelado geomorfológico e o processo de formação do solo, tanto no perfil longitudinal da paisagem como no latitudinal, influenciando também os processos de pedogênese.

A diferença de declividade influencia fortemente as perdas de solo, um processo causado pelo movimento da água e que origina,

consequentemente, perda da matéria orgânica, argila e outras partículas leves do solo (Oliveira et al., 2012). Quanto maior a inclinação do terreno, maior será o volume e a velocidade de escoamento superficial e menor será a capacidade de infiltração de água no solo. Esses processos aumentam a capacidade de transporte de partículas e as chances de desagregação do solo por ação de cisalhamento (Cogo et al., 2003).

A curvatura do perfil, é um atributo de segunda ordem em relação à altitude ou de primeira ordem em relação à declividade, refere-se ao caráter convexo do terreno (valores positivos) e caráter côncavo (valores negativos) quando analisada em perfil (Valeriano, 2008). A curvatura planar é um atributo de segunda ordem das curvas de nível, pode ser descrita como a variação da orientação de vertentes em uma determinada distância. Traduz-se no caráter divergente ou convergente dos fluxos de matéria sobre o terreno. Tem influência na acumulação da umidade e no fluxo de água superficial e subsuperficial do terreno (Valeriano, 2008). É a curvatura de uma superfície perpendicular à direção de inclinação mais acentuada e indica o acúmulo de água. Ambas as curvaturas são expressas normalmente em graus por metro ($^{\circ}/m$) (Valeriano, 2008).

A área de contribuição é uma variável extraída diretamente do MDE e amplamente utilizada na modelagem da hidrologia, erosão do solo e ambiente ecológico. A partir da obtenção da variável direção de fluxo é possível estimar a área de contribuição para cada célula, que é a área drenada a montante e está relacionada as áreas que ocorrem maior fluxo de água. Esse atributo é importante nos estudos de áreas de instabilidade, principalmente em áreas côncavas do relevo (Beven & Kirkby, 1979).

Os atributos topográficos secundários, ou compostos, envolvem combinações dos atributos primários e, geralmente, são índices geomorfométricos, tais como o índice de umidade topográfica (IUT), índice de posição topográfica (TPI), índice de poder de escoamento, entre outros. Essas variáveis ambientais podem alterar as características do solo e sua distribuição na paisagem, uma vez que à água se distribui sobre as vertentes e afeta a pedogênese do solo (Wilson e Gallant, 2000).

Um dos índices mais utilizados no MDS é o índice de umidade topográfica (IUT) que é definido como uma função da declividade e da área de

contribuição por unidade de largura ortogonal à direção do fluxo. Tem sido utilizado para modelar a distribuição espacial de zonas de saturação superficial e determinar o conteúdo de água nas paisagens. O índice mostra a extensão de acumulação de fluxo (Moore et al. 1991; Florinsky et al. 2002). Chagas (2006) avaliou as relações existentes entre material de origem e atributos do terreno de um modelo digital de elevação. Dentre as variáveis preditoras testadas, o índice TWI foi utilizado para separar os Gleissolos, que ocorrem em relevo plano, de outras classes de solos que ocorrem em áreas de relevo plano e suave ondulado.

O índice de posição topográfica (TPI), proposto por Weiss (2001), é derivado de um algoritmo que mede a diferença entre a elevação de um ponto central em relação à média da elevação dos pontos vizinhos (Jennes, 2006). Isso significa que quando os valores são positivos a célula tem um valor maior do que a sua vizinhança o que pode indicar um topo de uma colina. Os valores negativos representam os locais mais rebaixados em relação a sua vizinhança, caracterizando os vales ou depressões. Valores próximos a zero podem indicar a presença de áreas planas ou áreas da meia encosta.

2.6. Desagregação de polígonos de mapa

O trabalho de mapeamento de solos em muitos estudos baseia-se no resultado de análises fisiográficas (Epagri, 2004) e na relação existente entre as características da paisagem e os solos dominantes. As populações de solos que possuem características similares ocorrem dentro de cada unidade fisiográfica, as quais podem ser identificadas e delineadas em mapas que são elaborados por pedólogos experientes.

O conhecimento atualizado e detalhado da distribuição espacial dos solos e de suas propriedades é importante para o monitoramento, manejo e modelagem, assim como para o planejamento do uso da terra (Häring et al., 2012; Omuto et al., 2013). Na abordagem do mapeamento convencional de solos são realizadas incursões no campo para identificação das UTs, são feitos a descrição de perfil, a delimitação da área a ser estudada e a amostragem, gerando assim uma legenda preliminar de solos da área. Frente a riqueza desses dados é necessário extrair o máximo possível de informações, já que, para obtê-los são demandados tempo e recursos financeiros (Mendonça-Santos & Santos, 2003).

Uma alternativa para melhor utilização desses dados, é a desagregação espacial das UM de solos, através da qual se busca individualizar as classes de solos das UMs combinadas, obtendo mapas com maior resolução espacial e melhor aplicabilidade das informações desses trabalhos (Häring et al., 2012). Essa abordagem consiste na utilização de dados históricos de levantamentos, junto com técnicas de MDS e variáveis ambientais preditoras para desagregar as UMs combinadas e representar as classes de solos em um mapa digital, levando ao aprimoramento de mapas menos detalhados, para produzir mapas que diferenciem espacialmente classes de solos em nível de maior detalhamento (Subburayalu et al., 2014).

A desagregação de polígonos tem sido identificada como uma abordagem conceitual para adequar dados existentes de solos às demandas atuais, utilizando-se os conceitos pedológicos de formação do solo (Bui & Moran, 2001; Bui, 2004). Assim, com a desagregação das UMs combinadas, mapas de solos existentes e menos detalhados podem ser melhorados sem a necessidade de novos levantamentos, economizando tempo e recursos. A proposta é assegurar que o esforço para novos levantamentos associados com o mapeamento digital de solos seja minimizado.

Para obter um mapa com maior nível de detalhamento a partir da desagregação espacial é necessário realizar a individualização das classes de solos, que pode ser feita tanto mantendo as bordas dos polígonos (Bui & Moran, 2001; Häring et al., 2012), quanto sem manter os limites originais das UM (Nauman & Thompson, 2014; Subburayalu et al., 2014).

Recentemente, diversos estudos têm utilizado a desagregação para ampliar a utilidade dos mapas para a pesquisa. Li et al. (2012) empregaram a desagregação para melhorar o entendimento acerca das classes de aptidão agrícola. Os autores criaram mapas com UM simples e resolução espacial de 250 m. Segundo os autores, a metodologia pode ser usada para gerar mapas com maior resolução espacial e, assim, permitir o conhecimento mais preciso da capacidade de uso do solo e melhorar a modelagem ambiental e preditiva.

Nauman & Thompson (2014) utilizaram a abordagem para identificar áreas representativas em cada unidade do mapa para cada UM, com base em variáveis do terreno derivadas do modelo digital de elevação. Os autores consideram que a desagregação integra os temas conceituais de levantamentos

de solos históricos, em um método que possibilita refinar quantitativamente mapas de solos convencionais de forma a agregar valor e aumentar a aplicabilidade.

Odgers et al. (2014a) testaram o algoritmo *Through Resampled Classification Trees* (DSMART) para predizer a distribuição espacial das classes de solos desagregando as UM de um mapa convencional. O algoritmo desagregou todos os polígonos de forma relativamente rápida. Outros trabalhos testando essa técnica não tem demonstrado a capacidade de desagregar todas as UMs combinadas na área de interesse (Subbarayalu et al., 2014).

Sarmiento et al. (2015) propuseram a desagregação de mapas para individualizar as classes de solos em uma área de cultivo de videira. Os resultados foram considerados satisfatórios e possibilitaram melhorar o uso dos mapas de solos convencionais.

A metodologia para a desagregação de polígonos ainda não foi consolidada, já que, parte do trabalho é feito de forma manual, necessitando de adaptações. Portanto, é necessário estabelecer um protocolo que possa facilitar a entrada de dados e estabelecer um padrão para a metodologia.

3. ESTUDO 1 – PREDIÇÃO DA OCORRÊNCIA DE UNIDADES FISIAGRÁFICAS COM USO DE TRÊS ALGORITMOS DE ÁRVORES DE DECISÃO

3.1. Introdução

A crescente demanda por novas áreas de produção agropecuária exige atividades de levantamento e mapeamento de solos a nível detalhado. Segundo Santos et al. (2013), o Brasil conta com apenas 1,71% do território nacional com mapas de solos em escalas que efetivamente podem auxiliar na tomada de decisão, gestão e produção agrícola e pecuária (escala maior ou igual a 1:100.000).

Como os levantamentos de solos tradicionais nessas escalas são caros e demorados, há uma crescente necessidade de aplicar técnicas mais eficientes que venham a ser alternativas aos levantamentos convencionais de solos. Sendo assim, McBratney et al. (2003) formularam o conceito de mapeamento digital de solos (MDS) e propuseram essa técnica como alternativa para gerar mapas com maior resolução espacial de forma econômica e rápida. O MDS é fundamentado no uso de modelos de predição, que podem fornecer uma gama de informações sobre classes e propriedades dos solos. No entanto, é uma técnica que ainda não é totalmente aceita, necessitando de pesquisas que venham validar seus resultados (Abdel-Kader, 2011).

Nesse contexto, os levantamentos convencionais existentes adquirem importância como fonte de informações para subsidiar a calibração dos modelos de predição de classes e propriedades dos solos (Scull et al., 2005; Omuto et al., 2013). Diversos estudos no Brasil têm utilizado a técnica de MDS (Carvalho Junior et al., 2011; Giasson et al., 2011; Sarmiento et al., 2012, ten Caten et al., 2013). Estes autores têm comparado os métodos de predição da ocorrência de classes de solos em áreas não mapeadas.

Dentre as técnicas de classificação, os estudos têm evidenciado que o uso de árvores de decisão (AD) apresenta bons resultados no MDS e se destaca na discriminação de classes de solos, pois fornece um aumento da acurácia e permite classificar conjuntos distintos de dados (Giasson et al., 2011; Höfig et al., 2014; Adhikari et al., 2014), sendo um método bastante estudado e útil para resolver problemas relacionados ao levantamento e mapeamento de solos.

Existem diversos algoritmos que constroem uma AD, entre eles *RandomForest*, *REPTree*, *J48*, *SimpleChart* e *BFTree*. Os algoritmos utilizam testes estatísticos para selecionar os atributos que serão utilizados no nó raiz e demais nós para melhor classificar os dados. Todos os algoritmos têm um objetivo em comum, que é gerar a menor AD que classifique corretamente as variáveis ambientais preditoras, sem proporcionar perdas significativas de acurácia geral e índice Kappa. No entanto, nem todos os algoritmos são capazes de gerar AD com tamanho desejável para implementação no SIG.

Uma das diferenças entre os algoritmos de aprendizagem de AD é a função de escolha do atributo mais relevante. O cálculo do ganho para cada atributo e escolha do atributo que produz o maior ganho, até que todos os nós folha sejam puros, ou seja todos os exemplares tenham a mesma classificação, tem sido amplamente utilizado na construção de AD. Sendo assim, cada algoritmo pode apresentar resultados diferentes em relação a sua capacidade de predição das classes de solos, ocasionando variações na acurácia das regras de classificação e no índice Kappa.

Tendo em vista as necessidades elencadas de novos estudos que tenham por propósito selecionar e recomendar os melhores algoritmos para uso no MDS, este estudo teve como objetivo testar o modelo de árvores de decisão no MDS e realizar uma análise comparativa entre os algoritmos de mineração de dados *J48*, *Simple Chart* e *BFTree* em relação a capacidade de predição de unidades fisiográficas.

3.2. Materiais e Métodos

3.2.1. Área de estudo

A área de estudo foi a microbacia Córrego Tarumãzinho (Figura 1), situada na porção oeste do Estado de Santa Catarina (SC), localizando-se a 5 km da sede do município de Águas Frias, cujo acesso se faz pela rodovia SC-479. Ocupa uma área de aproximadamente 27,7 km² e apresenta cinco fases principais de relevo: plano, suave ondulado, ondulado, forte ondulado e montanhoso (Embrapa, 2004).

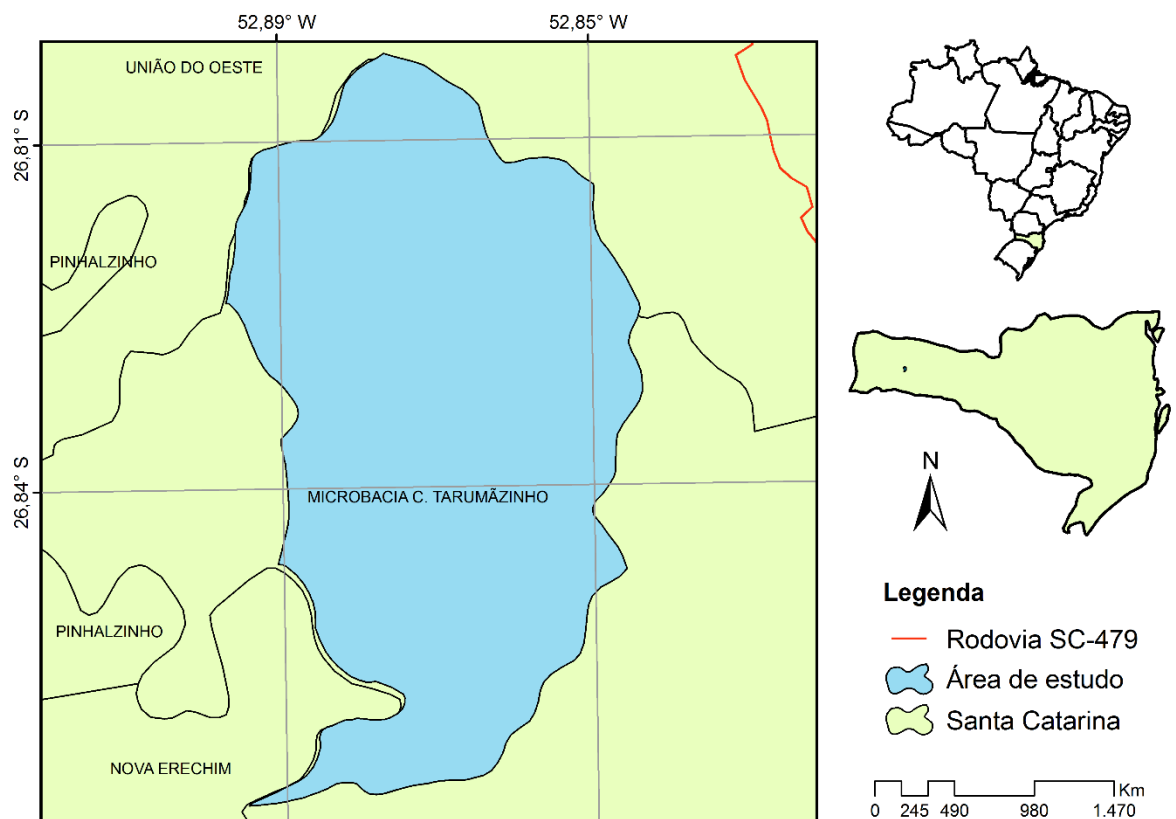


Figura 1. Localização da microbacia Córrego Tarumãzinho no oeste do Estado de Santa Catarina. Fonte: Epagri (2004).

O clima na região é classificado segundo Koppen como Cfa, caracterizado por ser constantemente úmido, com precipitação pluviométrica média anual de 1.725 mm e a temperatura média anual varia de 17,9 a 19,8°C. A temperatura média das máximas varia de 25,8 a 27,5°C, e das mínimas de 12,9 a 14,0 °C. A altitude varia de 283 metros a 660 metros. A microbacia dispõe de um mapa fisiográfico na escala de 1:25.000 e um relatório técnico, onde estão

descritos os perfis dos solos dominantes nas unidades fisiográficas e suas associações (Epagri, 2004).

Os solos que compõem a microbacia tem como principal material de origem as rochas basálticas do Grupo São Bento, da Formação Serra Geral. De acordo com Silva & Bortoluzzi (1987), o material de origem constitui-se por rochas vulcânicas em derrames basálticos de textura afanítica (rochas de granulação fina, arrefecimento rápido do magma à superfície e a cristalização pode não ser completa), amigdaloidal (textura de rocha vulcânica a sub-vulcânica contendo muitas amígdalas) no topo dos derrames, coloração cinza escura à negra, com intercalação de arenitos intertrapeanos.

Essas rochas basálticas são essencialmente constituídas por plagioclásios cálcicos e piroxênios, minerais com estabilidade bastante baixa, por este motivo alteram-se quase que totalmente a minerais argilosos com liberação de grande quantidade de óxidos, especialmente de ferro, dando origem a solos argilosos (Santa Catarina, 1986).

Uma característica marcante do basalto é a relativa facilidade de intemperização que, dependendo de outros fatores de formação do solo (principalmente relevo), dá origem a solos profundos a muito profundos. A parte superior do derrame (zona vesicular ou amigdalóide) é particularmente importante para formação dos solos. Quando essas cavidades são preenchidas com calcita, formam-se solos ricos em cálcio (Epagri, 2004).

A vegetação original direta ou indiretamente, desempenha papel importante sobre as características pedológicas da paisagem. A vegetação ocorrente na área é composta pelas seguintes formações vegetais: Floresta Estacional Decidual (Floresta Latifoliada do Alto Uruguai), localizada às margens do rio Uruguai estende-se pela microbacia até as altitudes de 500 a 600m e Floresta Ombrófila Mista, contempla o restante da microbacia nas altitudes acima de 500 metros (Santa Catarina, 1986).

3.2.2. Identificação das UFs e solos dominantes.

Na área em estudo foram identificados através de fotointerpretação seguida de mapeamento a campo por Epagri (2004), seis unidades fisiográficas (subpaisagens), sendo elas: Cume Erosional (CE), Encosta Coluvial Erosional (ECE), Encosta Erosional (EE), Encosta Erosional Coluvial (EEC), Encosta em

Patamar (EP), Fundo de Vale Coluvial Aluvial (FVCA). A ocorrência dessas UFs (Figura 2), suas características, comportamento e distribuição na área de estudo se relacionam diretamente com os fatores de formação dos solos, além de alguns fatores modificadores locais, dentre esses foram selecionados e utilizados para a análise fisiográfica: a declividade, a profundidade efetiva, a pedregosidade, a suscetibilidade à erosão, a fertilidade e a drenagem (Epagri, 2004).

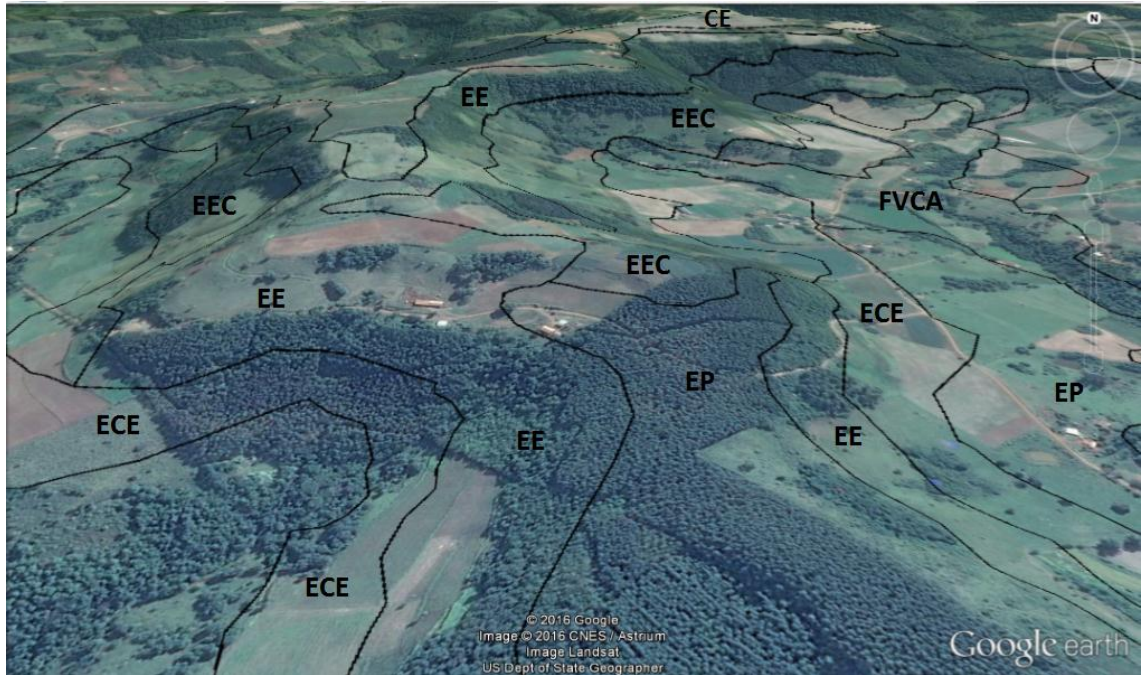


Figura 2. Visão 3D no Google Earth das unidades fisiográficas da microbacia Córrego Tarumãzinho.

O estudo de levantamento e mapeamento de solos nesta área baseou-se na análise fisiográfica e identificação dos solos dominantes, classificados de acordo com o Sistema Brasileiro de Classificação de Solos (Embrapa, 1999). Os solos dominantes foram identificados a partir de levantamento a campo, de acordo com as características morfológicas, físicas e químicas obtidas através de análises laboratoriais dos horizontes e sub-horizontes A e B e estão espacializados de acordo com a UF e representados originalmente em um mapa fisiográfico na escala aproximada de 1:25.000 (Figura 3).

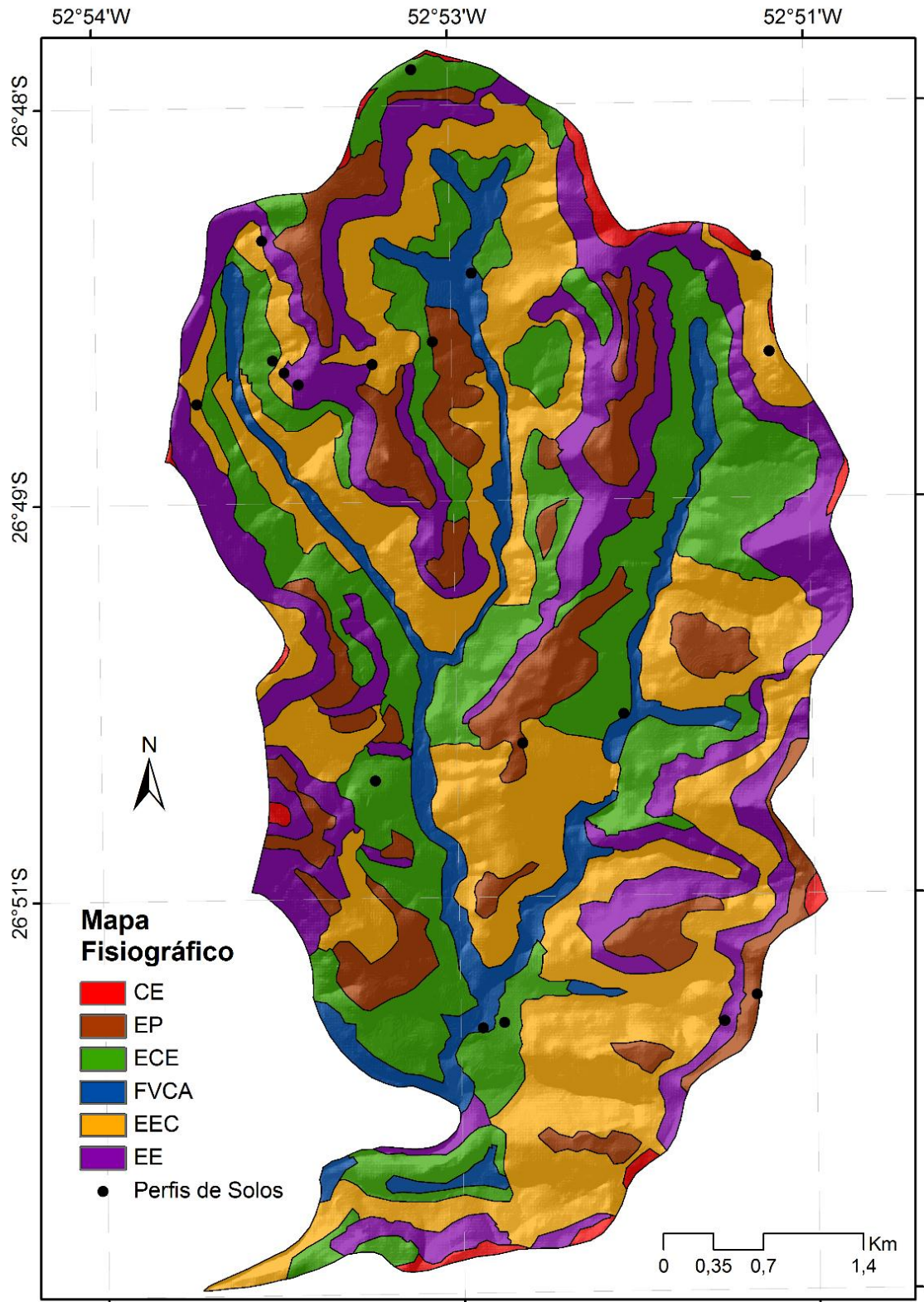


Figura 3. Mapa fisiográfico da microbacia Córrego Tarumãzinho. Fonte: Epagri (2004). Componentes: Cume Erosional (CE), Encosta Coluvial Erosional (ECE), Encosta Erosional (EE), Encosta Erosional Coluvial (EEC), Encosta em Patamar (EP), Fundo de Vale Coluvial Aluvial (FVCA).

A análise fisiográfica teve por princípio, o entendimento das condições de gênese e evolução das paisagens que apresentam estreita associação com os processos pedogênicos, o que possibilita o reconhecimento dos tipos de solos associados a cada UF (Botero, 1977; Goosen, 1971 e Villota, 1992 e 2005). Os solos dominantes e a área ocupada em cada UF estão descritos na Tabela 1. Nota-se, que há o predomínio de Cambissolos e Nitossolos em pelo menos cinco UF.

Tabela 1. Descrição das unidades fisiográficas e os solos ocorrentes na microbacia Córrego Tarumãzinho. Fonte: Epagri (2004).

Unidades Fisiográficas	Solos correspondentes	Inclusões	Área	
			(%)	(ha)
CE	Nitossolo Háplico + Nitossolo Vermelho	Cambissolo	1,4	38
ECE	Cambissolo Háplico + Nitossolo Háplico + Nitossolo Vermelho	-	24,3	669
EE	Neossolo Litólico	Cambissolo	21,1	580
EEC	Cambissolo Háplico	Nitossolos	32,7	900
EP	Cambissolo Háplico	-	12,3	339
FVCA	Cambissolo Háplico	Gleissolos	8,2	226
Total			100,0	2752

3.2.3. Metodologia

Para execução e avaliação da técnica de MDS, foi necessário criar uma base de dados espacial para alimentar o Sistema de Informações Geográficas (SIG). Para construir a base de dados foram selecionadas e relacionadas diversas variáveis, tais como, o MDE, as variáveis ambientais preditoras (derivadas do MDE), as unidades fisiográficas (variável resposta) e os pontos amostrais dos perfis georreferenciados.

Os dados selecionados foram utilizados como referência sobre a distribuição das UFs a fim de calibrar modelos matemáticos e estatísticos que descreveram a ocorrência das UFs na paisagem a partir da relação com variáveis ambientais, levando em consideração que cada UF está relacionada com a ocorrência de um ou mais tipos de solos.

Estes modelos foram utilizados para produzir o mapa digital fisiográfico (MDF), contendo informações sobre a distribuição espacial das UFs. A partir do MDT com resolução espacial de 5 m, foram derivadas as variáveis ambientais preditoras com uso do software System for Automated Geoscientific Analyses (SAGA GIS). Foram obtidas doze variáveis ambientais, além da própria variável elevação, sendo elas: a) declividade, b) curvatura do perfil, c) curvatura planar, d) curvatura (combinação entre curvatura planar e curvatura do perfil), e) direção de fluxo, f) acúmulo de fluxo, g) índice de umidade topográfica (Beven & Kirkby, 1979). Além destas, foram utilizadas as variáveis, área de contribuição, insolação difusa, distância diagonal da drenagem e geoforma, sendo que está última foi gerada com auxílio de um conjunto de programas chamado LandMapr toolkit[©] MacMillan (2003).

As variáveis ambientais preditoras foram escolhidas porque podem representar alterações sobre os fatores de formação do solo e, portanto, são úteis na predição de ocorrência do mapa fisiográfico. Como procedimento inicial do software, foi necessário realizar o pré-processamento dos dados a fim de tornar as bases de dados compatíveis com o formato da ferramenta. O pré-processamento da matriz de dados constitui na retirada de inconsistências, retiradas de informações que não contribuem para o modelo gerado.

No processo de amostragem, foi utilizada a ferramenta *Create Random Points* encontrada no ArcGIS 9.3[®]. Com o uso da ferramenta foi realizada a amostragem sobre o mapa fisiográfico, de forma estratificada aleatória, para que a distribuição dos pontos seja proporcional ao número de polígonos e área ocupada por cada UF. Cada ponto foi utilizado para obter informações das 12 variáveis preditoras, esses dados formaram uma tabela com 7084 linhas correspondentes aos pontos amostrais e 12 colunas referentes às variáveis. Para essa amostragem, foi usado o módulo *Sample* encontrada na extensão *Spatial Analyst* em ambiente ArcGIS 9.3[®].

Depois de criados os pontos com os dados amostrais no ArcGIS 9.3[®], os mesmos foram exportados e convertidos em um arquivo delimitado por vírgula (CSV format - *Comma Separated Values*) para adequação e uso no software Weka 3.6.6 (*Waikato Environment for Knowledge Analysis*) e usados no treinamento dos algoritmos de árvores de decisão. O pacote Weka consiste de uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Pode ser usado para aplicar métodos de aprendizado a um conjunto de dados e analisar a saída para extrair informações a partir dos dados de entrada (Hall et al., 2009).

O Weka usa arquivos de dados de treinamento onde devem ser explicitadas quais variáveis são permitidas para uma relação específica, bem como o tipo de dado de cada variável, isto é, nominal ou valor numérico. O software pode detectar padrões em dados que podem ser explorados mediante regras.

Com a ferramenta *Select Attributes*, no Weka foram verificadas as variáveis que possuem maior capacidade para discriminação entre as UFs. Foram testados os algoritmos de árvores de decisão *J48* (Quinlan, 1993), *SimpleChart* (Breiman et al., 1984) e *BFTree* (Shi 2007), com número de elementos no nó final (M) igual a 5 (M5), possibilitando gerar AD menores e menos complexas sem reduzir demasiadamente a acurácia dos modelos e a quantidade das UFs preditas (Hall et al., 2009).

Os mapas de predição de ocorrência das UFs foram gerados a partir da implementação das regras de classificação proveniente dos modelos de AD no ArcGIS 9.3[®] utilizando a função *Raster Calculator* e seus operadores lógicos nível a nível da AD até chegar aos nós finais e gerar camadas com a delimitação de cada UF do mapa fisiográfico. Para implementação da árvore em ambiente ArcGIS as regras de classificação foram ajustadas a partir de um algoritmo elaborado em C# (C Sharp) de forma a serem diretamente utilizadas em programas SIG.

A concordância dos mapas produzidos com o mapa fisiográfico foi avaliado com o uso de matrizes de erro (Congalton, 1991), sendo que a referência foi o mapa fisiográfico da microbacia.

Foi calculado a concordância pixel a pixel, considerando: acurácia geral (AG), que é a proporção dos pixels corretamente classificados em relação

ao número total de pixels; acurácia do mapeador (AM), que trata da proporção que certa UF foi mapeada corretamente; acurácia do usuário (AU), que revela a certeza que o usuário tem, ao analisar o mapa, que um ponto específico, em uma determinada UF está mapeada corretamente e o valor do índice Kappa, obtido a partir da matriz de confusão. (Giasson et al., 2011). O índice Kappa utilizado para análise da classificação é fornecido pelo aplicativo WEKA após a fase de validação.

3.3. Resultados e discussão

A partir do mapa fisiográfico, foram gerados diferentes mapas de variáveis ambientais e a ferramenta *Select Attributes*, no Weka, indicou que as variáveis mais representativas para prever a ocorrência das UFs foram a distância diagonal da drenagem, a área de contribuição, a altitude e a declividade, as quais têm sido comumente relatadas como importantes variáveis ambientais para o MDS (Crivelenti et al., 2009; ten Caten et al., 2012; Giasson et al., 2013; Teske et al., 2015).

Quanto à influência de cada uma das variáveis ambientais preditoras, Morgan & Bui (2002) verificaram que os atributos do terreno mais recorrentes quanto da utilização de árvores de decisão foram área de contribuição, distância diagonal da drenagem, aspecto, declividade e curvatura, sendo que a área de contribuição está relacionada à disponibilidade de água para os processos ambientais.

A grande presença de cursos d'água na microbacia faz com que a distância diagonal da drenagem não seja elevada, com 67% da área apresentando distâncias muito pequenas (8 a 20 m) e 22% distâncias pequenas (20 a 45 m). As distâncias médias (45 a 70 m) também possuem representatividade na área de estudo, ocupando aproximadamente 9 % desta. Em contrapartida, distâncias grandes (70 a 150 m) ficam limitadas às áreas de maior altitude, com baixa representatividade local e representam cerca de 2% da microbacia (Figura 4).

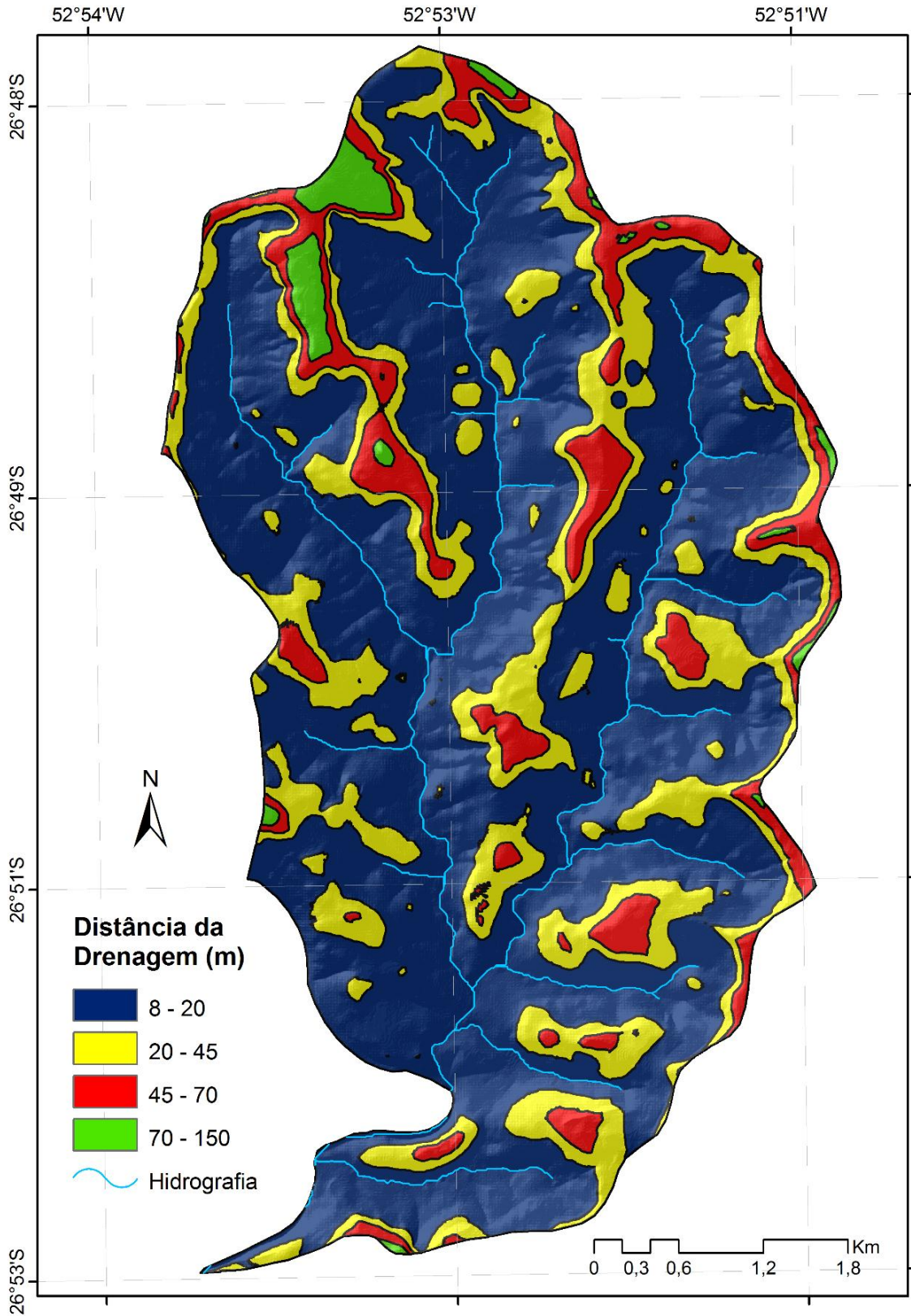


Figura 4. Distância diagonal da drenagem da microbacia Córrego Tarumãzinho.

Quanto à área de contribuição (Figura 5), há predominância das classes muito baixa (menor que 12.000 m²) e baixa (12.000 a 150.000 m²), compreendendo 99 % da microbacia, sendo que algumas porções da microbacia apresentam grandes áreas de contribuição (maior que 150.000 m²), porém representam apenas 1 % da área, segundo a classificação de Quinn et al. (1991).

A altitude média da microbacia Córrego Tarumãzinho é de 471 metros acima do nível do mar e a altitude máxima encontrada na região é de 660 metros, sendo que as maiores altitudes se concentram no limite norte da UF Cume Erosional. As menores altitudes estão em torno de 283 metros e se localizam ao longo da UF Fundo de Vale Coluvial Aluvial (Figura 6).

Outra variável ambiental importante na predição das UFs foi a declividade (Figura 7), sendo que a microbacia apresenta em sua maioria áreas com relevo ondulado a forte ondulado, compreendendo 38 e 43 %, respectivamente, da área de estudo. Áreas com relevo plano e suave ondulado são minoria, juntas compreendem 10 % da microbacia, havendo, porém, 9 % de áreas com relevo montanhoso, segundo a classificação de Gallant & Wilson (2000).

As variáveis distância da drenagem e declividade estão incluídas entre as mais importantes porque são associadas ao acúmulo de água e aos fluxos hídricos, que têm grande influência na formação dos solos. (Crivelenti et al., 2009).

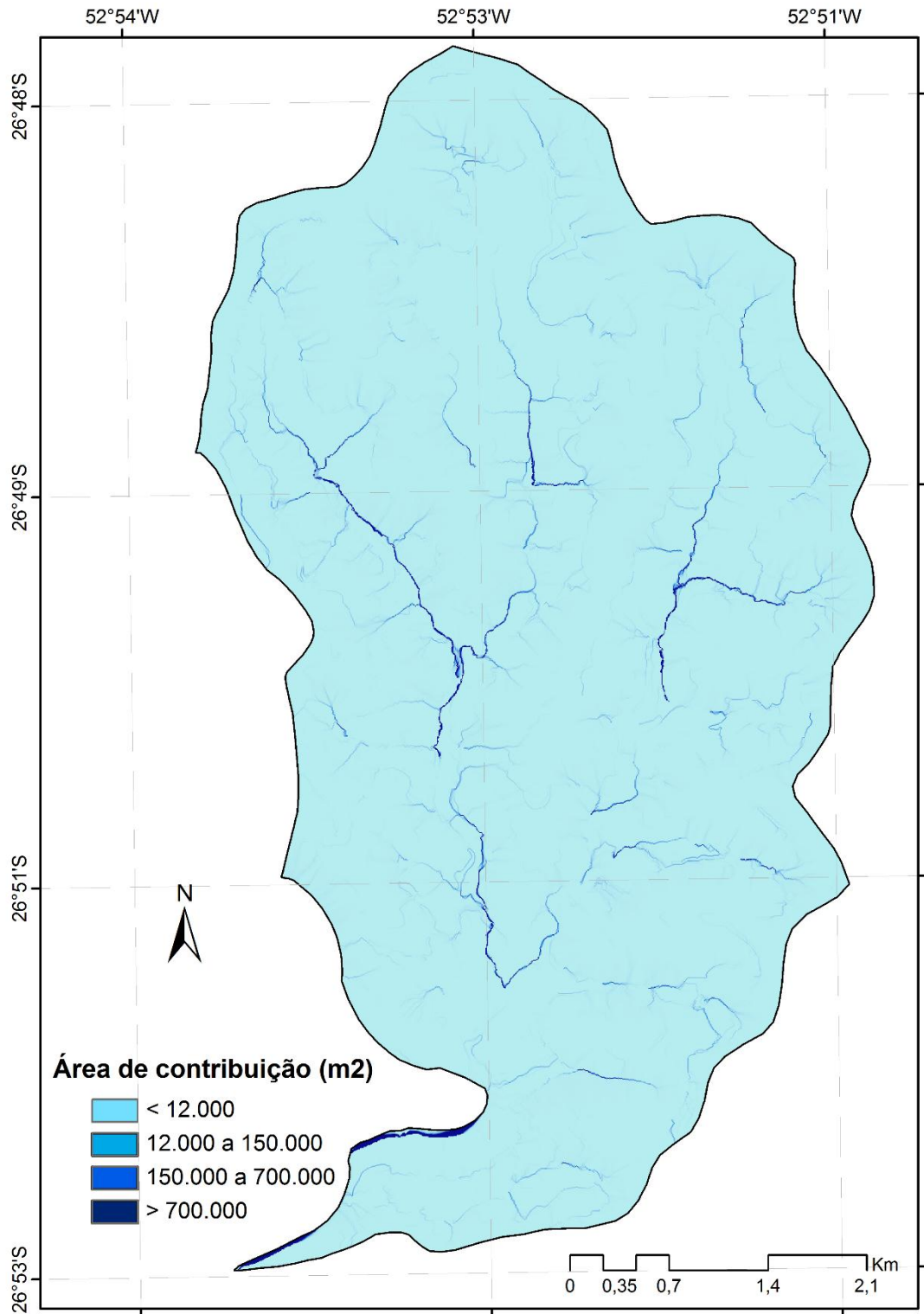


Figura 5. Área de contribuição da microbacia Córrego Tarumãzinho.

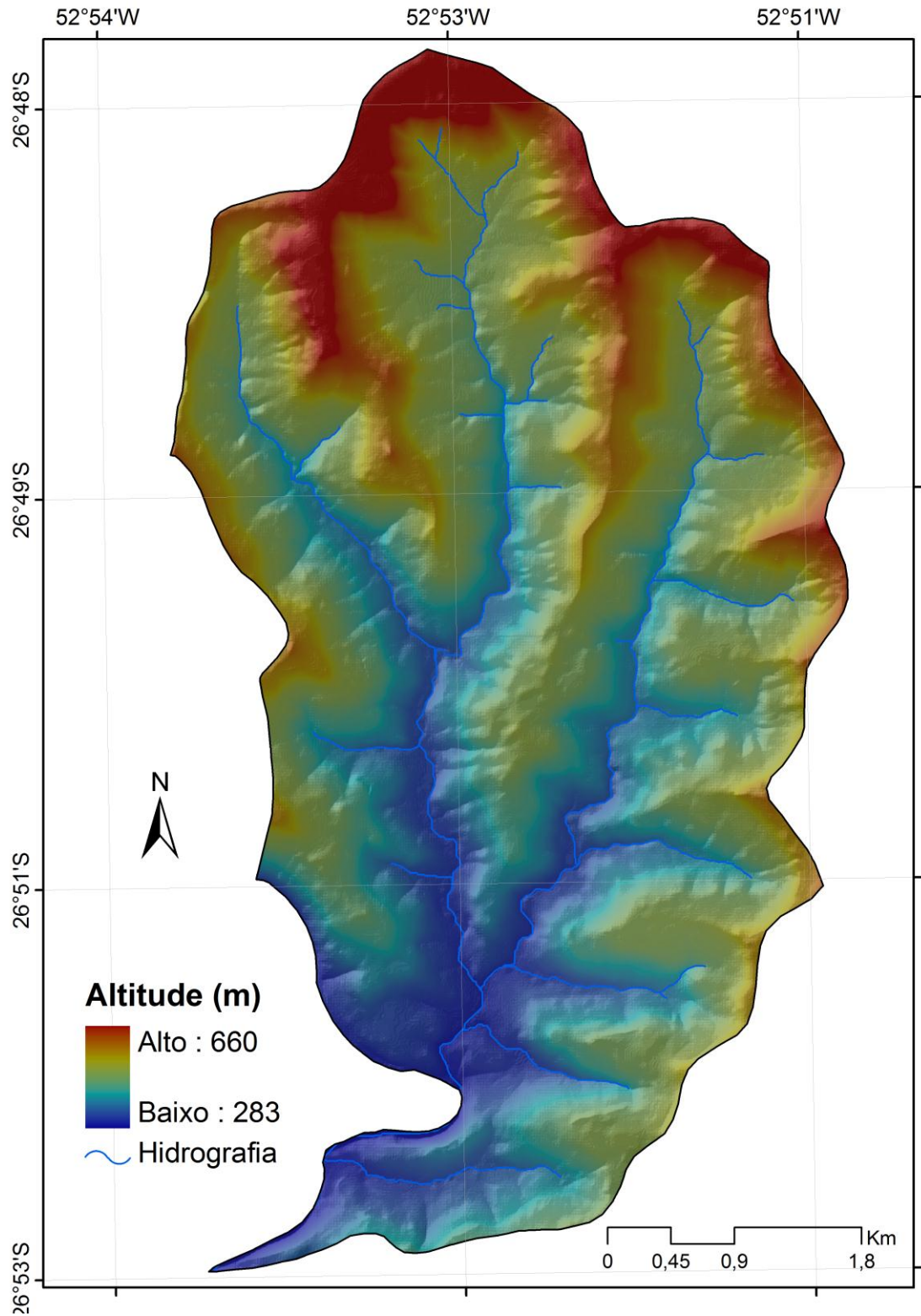


Figura 6. Altitude da microbacia Córrego Tarumãzinho.

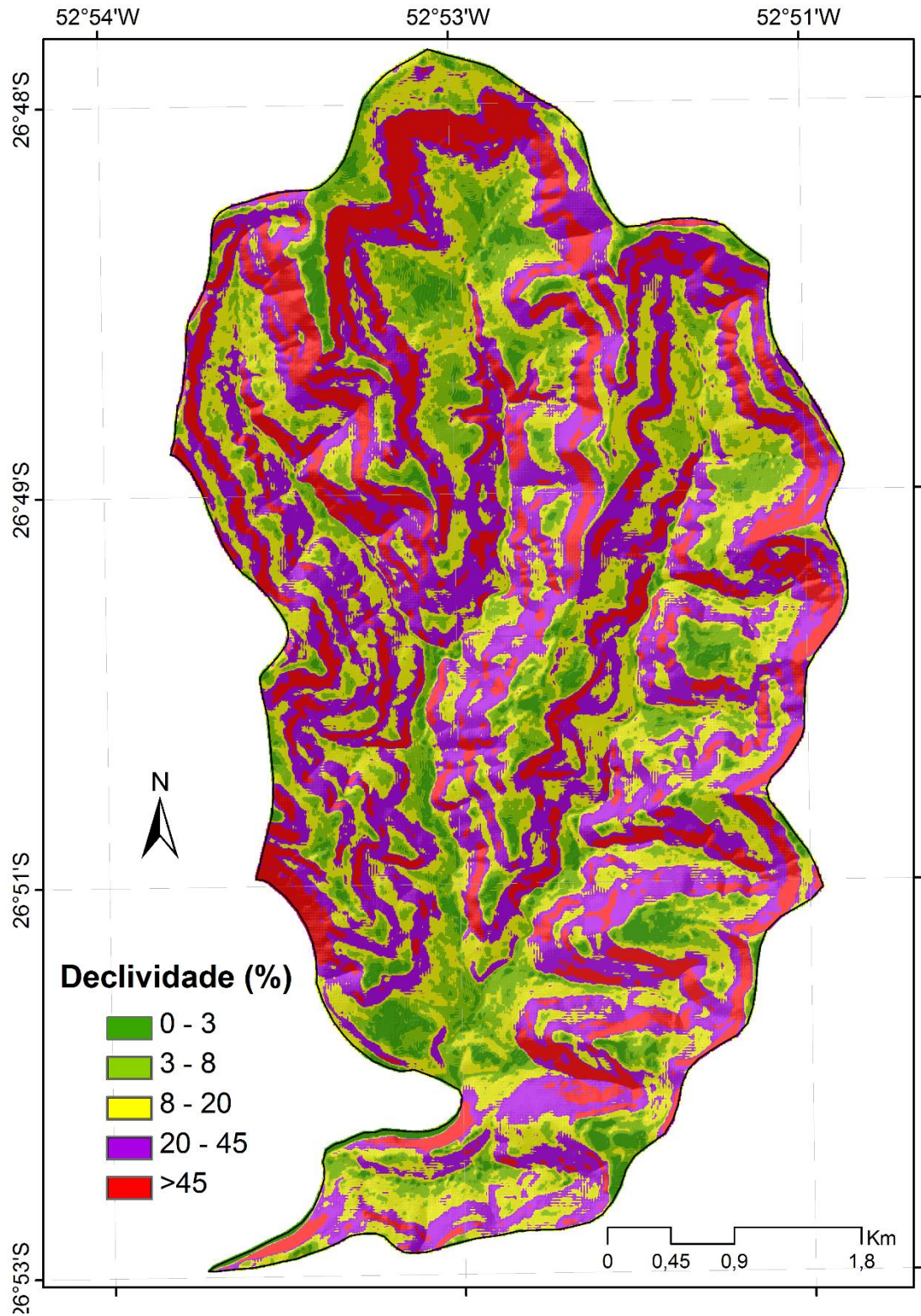


Figura 7. Declividade do terreno na microbacia Córrego Tarumãzinho.

3.3.1. Algoritmos de Árvores de Decisão

Por meio de comparação visual, é possível verificar a existência de correlação entre o mapa predito e o utilizado como referência (Figuras 3 e 8). Essa concordância, embora não represente um resultado de acurácia, permite afirmar, que há uma correspondência entre os dois mapas. Essa correspondência revela que o modelo foi eficiente na tentativa de representar as UFs preditas o mais fidedigno possível as UFs do mapa fisiográfico.

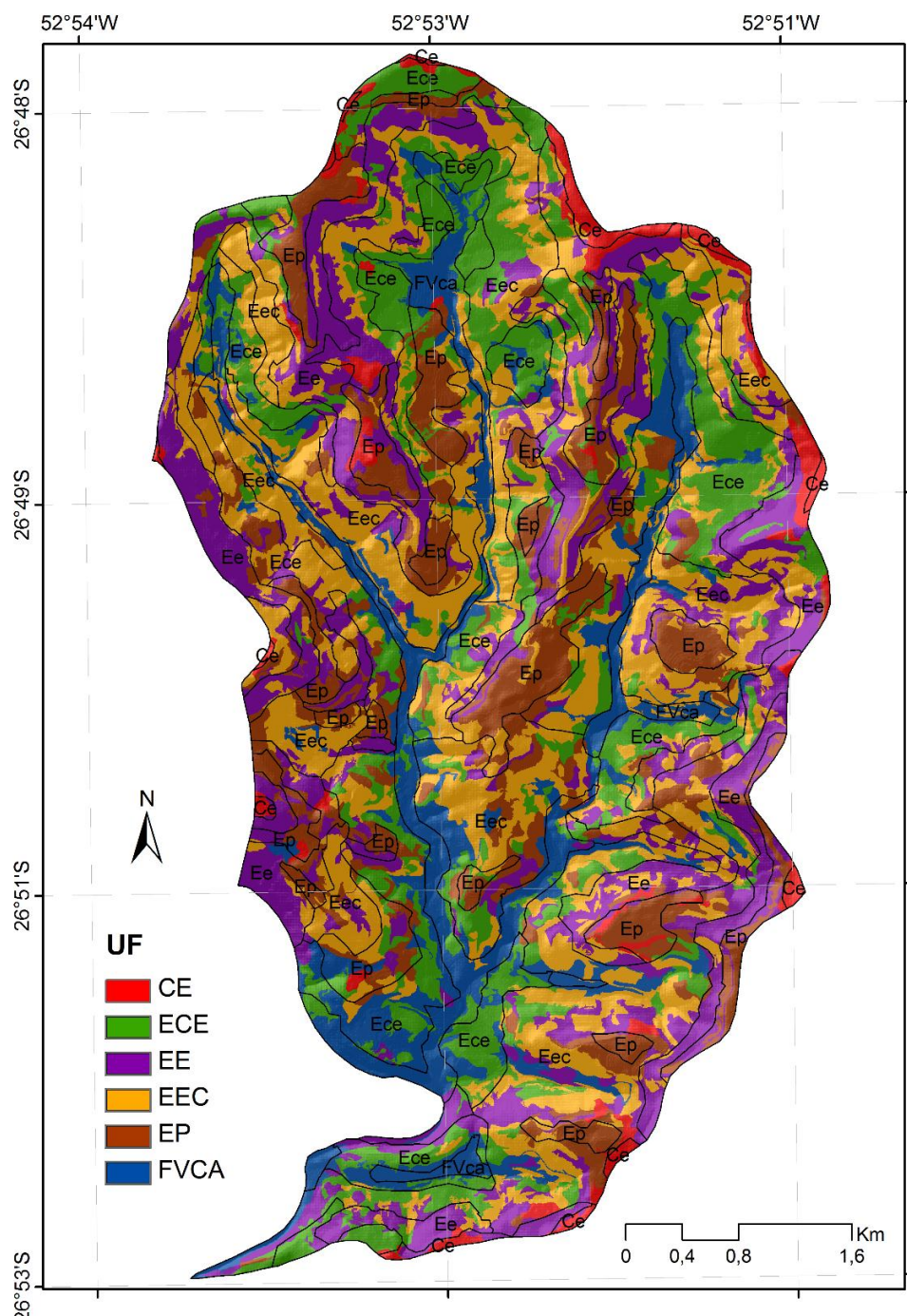


Figura 8. Mapa fisiográfico predito produzido pelo algoritmo *J48* e mapa convencional da microbacia Córrego Tarumãzinho.

Ao analisar a matriz de erros entre o mapa fisiográfico e o mapa predito com o algoritmo *J48* (Tabela 2), verificou-se que a AG, que representa a concordância entre o mapa predito realizado com o modelo de AD e o mapa fisiográfico, foi de 65,9%. O índice Kappa global que considera todas as UFs classificadas corretamente foi de 0,57 indicando qualidade moderada, segundo o critério de Landis & Koch (1977). Esses resultados estão de acordo com Giasson et al. (2013) que testaram três diferentes MDEs e cinco algoritmos árvores de decisão para a reprodução de mapa semidetalhado de solos. Os autores concluíram que o algoritmo *J48*, seguido do *Simple Chart* e *BFTree* foram os algoritmos que produziram mapas de solo com maior concordância com o mapa de referência.

Tabela 2. Matriz de erro do Algoritmo *J48*, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.

		Mapa Fisiográfico							Total	AU %
		UM	CE	ECE	EE	EEC	EP	FVCA		
Mapa predito	CE	900	18	26	33	36	0	1013	85,2	
	ECE	33	615	115	214	77	123	1177	44,4	
	EE	43	113	739	195	115	20	1225	54,1	
	EEC	38	173	197	696	166	104	1374	48,0	
	EP	51	71	118	146	669	14	1069	65,1	
	FVCA	1	110	21	93	15	986	1226	73,8	
	Total	1066	1100	1216	1377	1078	1247	7084	61,8	
AM %	88,8	52,3	60,3	50,7	62,6	80,4				

Acurácia Geral = 65,9; Kappa = 57,9%; Tamanho da árvore: 508 folhas

Ainda analisando a matriz do algoritmo *J48*, verifica-se que a unidade fisiográfica mapeada com maior acurácia do mapeador (AM) foi a unidade CE com 88,8%. Isso ocorreu, provavelmente, pelo sistema de amostragem utilizado, de forma estratificada aleatória, que considerou a distribuição de pontos em

todos os polígonos. Por ser a unidade de menor extensão, com cerca de 38 hectares, os pontos foram melhores distribuídos em CE permitido maior classificação dos pixels para essa unidade. Hengl et al. (2003), testaram diferentes tipos de amostragem e concluíram que a amostragem estratificada foi o método mais apropriado para prever a ocorrência de classes de solos em área de menor extensão.

Em contrapartida, a menor concordância de AM foi verificada na unidade fisiográfica EEC com 50,7%, este percentual representa a área que foi mapeada corretamente. A unidade EEC foi a mais representativa em termos de extensão da área, compreendendo 900 hectares. Já na AU, observou-se que a correspondência para a UF Cume Erosional se manteve elevado (85,2%), mostrando que essa UF teve alto índice de pontos classificados corretamente no mapa fisiográfico predito e a probabilidade de que essa unidade corresponda a verdade do mapa fisiográfico é bastante alta. A menor AU foi encontrada na UF Encosta Coluvial Erosional (44,4%), mostrando baixa correlação entre o mapa fisiográfico predito e o mapa fisiográfico (Tabela 2).

Os valores de AG e AU podem variar consideravelmente. Por exemplo, Crivellini et al. (2009), trabalhando com a folha Dois Córregos/SP, chegou a AG de 54,2%. Sarmiento et al. (2012) testaram um MDE com cinco metros de resolução e uma base cartográfica digital em escala 1:5.000 para predição de classes de solos no Vale dos Vinhedos/RS e obtiveram AG de 71%. Essa variação pode estar relacionada com a técnica utilizada, a qualidade das variáveis ambientais testadas, o tamanho da área de estudo e o sistema de amostragem utilizado.

Na Tabela 3, estão representados os percentuais de registros corretamente classificados pelo algoritmo *Simple Chart*. A acurácia geral (AG) foi de 64,5% e índice Kappa de 0,56, esses resultados são levemente inferiores aos apresentados pelo algoritmo *J48*. Em relação ao índice Kappa, não existe um valor específico a partir do qual se deva considerar o índice como adequado, encontram-se na literatura algumas sugestões que orientam normalmente esta decisão, sugerindo que valores de Kappa entre 0,4 e 0,6 são satisfatórios ou bons (Landis & Koch, 1977; Fonseca et al., 2007).

Tabela 3. Matriz de erro do Algoritmo *SimpleChart*, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.

		Mapa Fisiográfico								
		UM	CE	ECE	EE	EEC	EP	FVCA	Total	AU %
Mapa predito	CE	868	32	32	36	45	0	1013	83,3	
	ECE	58	580	106	190	99	144	1177	37,7	
	EE	47	79	730	229	115	25	1225	46,0	
	EEC	49	156	217	692	167	93	1374	42,0	
	EP	71	69	149	132	633	15	1069	55,2	
	FVCA	1	99	19	84	11	1012	1226	70,3	
	Total	1094	1015	1253	1363	1070	1289	7084	55,8	
AM %	85,7	49,3	59,6	50,4	59,2	82,5				

Acurácia Geral = 64,5; Kappa = 56,4%; Tamanho da árvore de decisão: 304 folhas

Sun et al. (2011) empregaram árvores de decisão e análise linear discriminante em estudo com dados legados de solos em zonas montanhosas e obtiveram índice Kappa de 0,63 gerando mapa final de classes de solos com boa precisão. Esses resultados confirmam que o modelo de AD é eficiente na discriminação de classes de solos em áreas de relevo muito dissecado.

Considerando-se uma avaliação visual do mapa gerado a partir das variáveis ambientais, verifica-se, no geral, uma significativa semelhança entre os mesmos (Figuras 3 e 9). Há, visualmente, uma maior concordância para as UFs menos representativas em termos de extensão CE e FVca, situadas respectivamente em áreas de relevo predominantemente plano, suave ondulado com ocorrência isolada de pequenas lombas.

Segundo Gonzalez et al. (2008), em áreas de pequenas extensões o relevo exerce maior influência sobre as características dos solos, dessa forma as variações em razão do clima são muito pequenas. Assim, classes que apresentam alta correlação com as variações do relevo se tornam mais fáceis de serem separadas através do MDS e, conseqüentemente, áreas com pequenas extensões apresentam maior AM.

A AM para as unidades fisiográficas CE e FVca foi de 85,7% e 82,5%, respectivamente, a semelhança entre os mapas pode ser facilmente observada

no mapa de concordância entre o mapa fisiográfico predito elaborado a partir de variáveis ambientais e o mapa fisiográfico, sendo que o mapa predito apresenta maior detalhe cartográfico (Figuras 3 e 9). Nas áreas mais dissecadas, que incluem as unidades fisiográficas ECE, EE, EEC e EP, os mapas são mais discordantes. Essa discordância pode ser explicada pela maior complexidade da distribuição espacial dos solos nessas unidades fisiográficas, em razão da geologia e relevo mais acidentado (Filho, 2010).

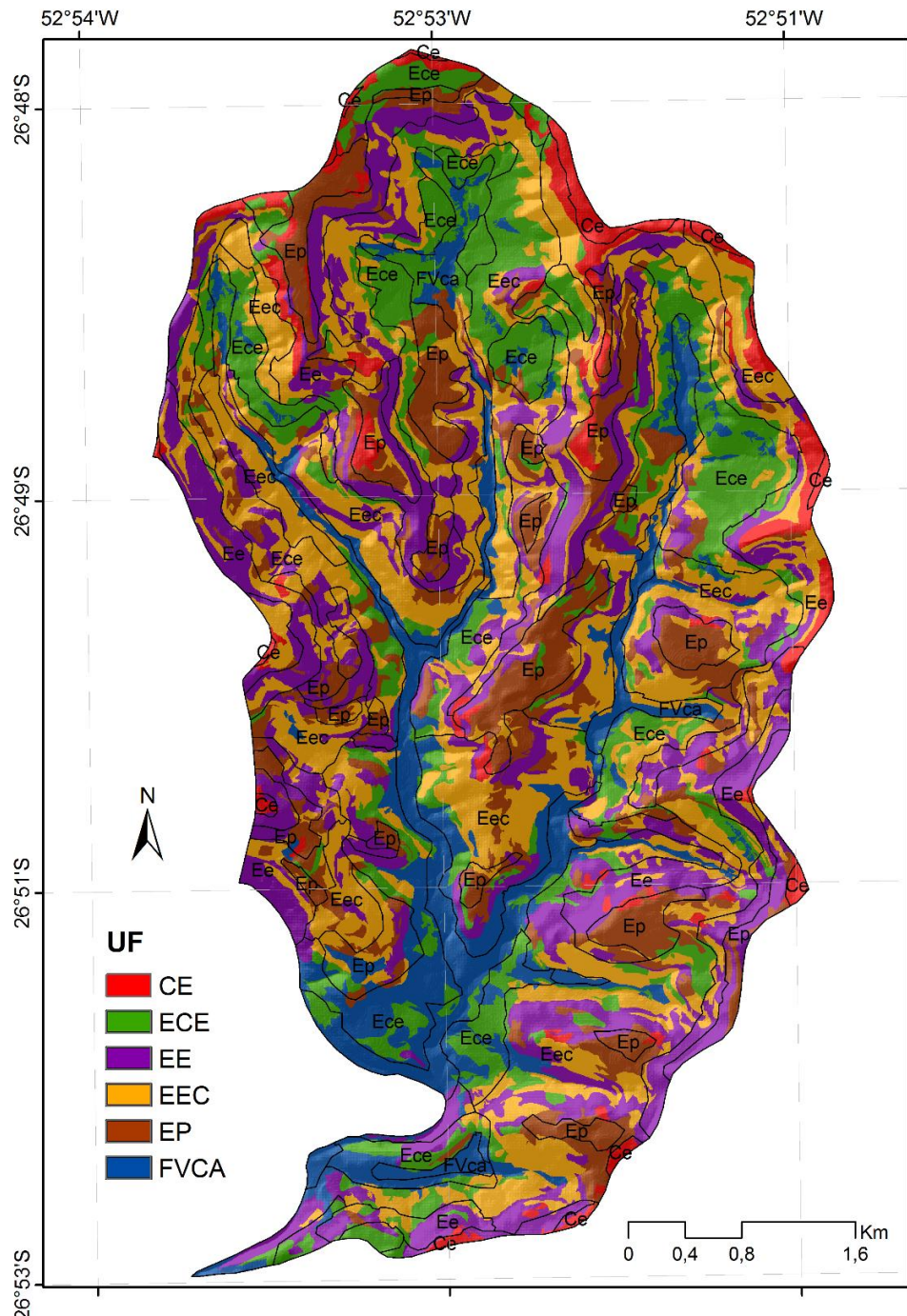


Figura 9. Mapa fisiográfico predito produzido pelo algoritmo *Simple Chart* e mapa convencional da microbacia Córrego Tarumãzinho.

A maior AU foi encontrada na unidade fisiográfica CE (83,3%), apesar de ser uma unidade de pequena extensão, os pixels foram melhores classificados em comparação com as unidades fisiográficas de maior extensão. Sato et al. (2013), realizaram uma análise comparativa entre doze algoritmos de mineração de dados e o algoritmo *Simple Chart* se diferenciou dos demais algoritmos devido ao tamanho da árvore, pois o número de folhas é menor em relação aos demais algoritmos, mostrando que esse método necessitou de um menor número de testes lógicos para a determinação das classes, o que facilitou sua implementação em ambiente SIG.

O algoritmo *BFTree* obteve valor de AG de 62,9% (Tabela 4). No geral, a árvore de decisão gerada com *BFTree* foi mais simples, com menor tamanho, expressos pelo menor número folhas. Isso é uma vantagem do modelo de AD, visto que, a implementação das regras dos modelos de árvores de decisão é realizada manualmente para a confecção do mapa de solos em SIG e que a geração de árvores menores facilita a sua implementação e reduz a chance de introdução de erros (Giasson et al., 2013).

Tabela 4. Matriz de erro do Algoritmo *BFTree*, comparando o mapa fisiográfico da microbacia Córrego Tarumãzinho com o mapa fisiográfico predito.

		Mapa Fisiográfico							Total	AU %
		Um	CE	ECE	EE	EEC	EP	FVCA		
Mapa predito	CE	859	30	55	31	38	0	1013	84,0	
	ECE	57	557	125	206	81	151	1177	35,7	
	EE	41	83	753	222	102	24	1225	47,0	
	EEC	49	148	262	692	134	89	1374	42,2	
	EP	69	63	194	175	554	14	1069	52,4	
	FVCA	1	93	21	94	12	1005	1226	71,3	
	Total	1076	974	1410	1420	921	1283	7084	55,4	
AM %	84,8	47,3	61,5	50,4	51,8	82,0				

Acurácia Geral = 62,9; Kappa = 54,7%; Tamanho da árvore de decisão: 299 folhas

Em relação a AM que indica a proporção que cada UF foi classificada corretamente, a UF mapeada com maior AM foi a unidade CE com 84,8% e o menor valor de AM 47,3%, foi observado na unidade ECE. Como a área de estudo apresenta polígonos pequenos de UFs, a escolha do sistema de amostragem estratificada, certamente contribuiu para que as unidades de menor extensão não sofressem alterações na acurácia. Como constatado em diversos trabalhos com MDS, onde a predição de ocorrência de UMs pouco representativas é prejudicada quando utilizada a amostragem aleatória (Giasson et al., 2011; Ten Caten et al., 2012). Sendo assim, a amostragem estratificada foi útil na predição de unidades fisiográficas em áreas que apresentam polígonos pequenos, já que a distribuição é proporcional ao número e a área ocupada por cada polígono. Dessa forma, todas as UFs foram classificadas pelos algoritmos (Tabela 5).

O coeficiente Kappa de 0,54 obtido para o Algoritmo *BFTree* foi considerado bom (Tabela 4), que segundo Fonseca et al. (2007) indica uma concordância média dos resultados preditos com os observados. Em estudo com algoritmos de AD aplicado a um conjunto de variáveis ambientais semelhantes as utilizadas no presente trabalho, Bui et al. (1999) obtiveram coeficiente Kappa para as classes individuais de 0,23 a 0,89, e coeficiente Kappa geral do mapa de 0,64, representando semelhança substancial dos resultados do modelo.

As unidades fisiográficas ECE e EEC foram as unidades que apresentaram as maiores diferenças em relação ao mapa de referência, confirmando os resultados da acurácia (Figuras 3 e 10). Essas diferenças ocorreram porque o *BFTree* confundiu a classificação entre as duas UFs. Esse confundimento ocorreu, provavelmente, em razão da unidade ECE ser formada no final da unidade EEC através da deposição gradual de sedimentos de solos desprendidos por força da erosão pluvial e erosão laminar, sendo arrastados encosta abaixo por incidência da gravidade (Epagri, 2004). Já, a unidade EEC, cujo comportamento e forma são determinados pelas diferentes declividades, alternância de áreas erosionais e coluviais associadas aos comprimentos de rampa variáveis, é formada logo abaixo das encostas erosionais e/ou das encostas em patamar.

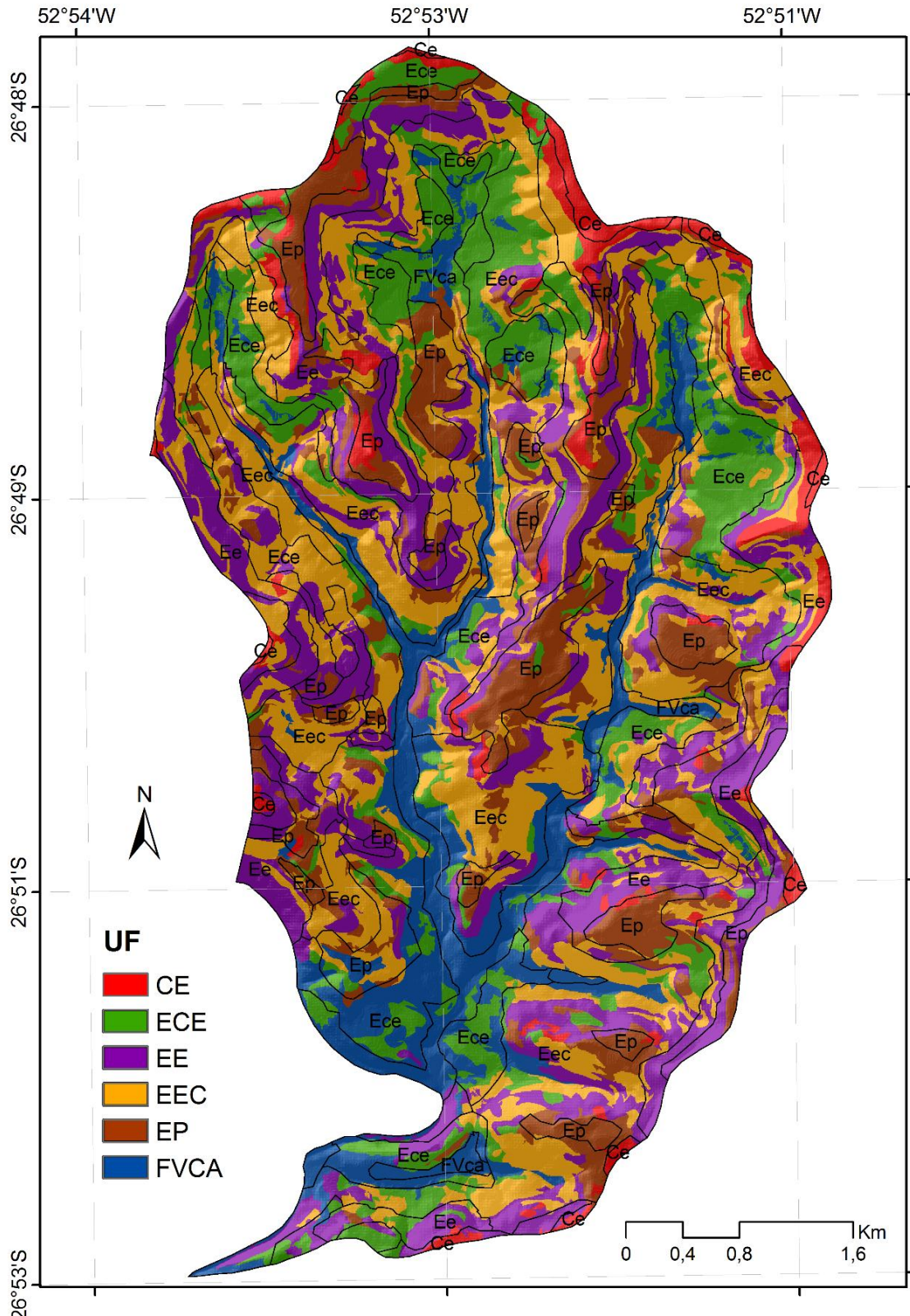


Figura 10. Mapa fisiográfico predito produzido pelo algoritmo *BFTree* e mapa convencional da microbacia Córrego Tarumãzinho.

Na Tabela 5, compara-se os três algoritmos em relação à acurácia geral, índice Kappa, tamanho da árvore de decisão e unidades fisiográficas que

os algoritmos foram capazes de prever. Verifica-se que o algoritmo *J48* foi superior quanto a acurácia geral (65,9%) e índice Kappa (0,57). No entanto, em relação ao tamanho da árvore de decisão o algoritmo *BFTree* apresenta menor tamanho de árvore, ou seja, a chance de erro na implementação das regras para a confecção do mapa fisiográfico em SIG foi reduzida (Giasson et al., 2013). No geral, todos os algoritmos conseguiram prever o total de 6 UF e os resultados comprovam que os mesmos também são eficientes para predição de unidades fisiográficas.

Tabela 5. Comparação dos resultados de acurácia, índice Kappa, tamanho da árvore de decisão e unidades fisiográficas preditas pelos três algoritmos.

Algoritmo	AG	K	t	UF
J48	65,9	0,57	508	6
Simple Chart	64,5	0,56	304	6
BFTree	62,9	0,54	299	6
Média	64,4	0,6	370	6

AG = acurácia geral; K = índice Kappa; t = tamanho da árvore de decisão; UF = unidades fisiográficas que o algoritmo foi capaz de prever

Os resultados deste estudo apresentam bons percentuais de AGs e AUs em comparação a outros estudos que testaram o modelo de AD. Partindo desse pressuposto, verifica-se que os resultados obtidos contribuem para o MDS no desenvolvimento de técnicas que auxiliem no mapeamento digital de unidades fisiográficas, mostrando-se eficiente nesse tipo de abordagem. No entanto, como as demais técnicas, necessita de aperfeiçoamento na rotina de trabalho no SIG, assim como o desenvolvimento de um protocolo, para que se obtenha valores de acurácia desejáveis.

3.4. Conclusões

As acurácias obtidas usando os três algoritmos foram semelhantes, sendo que o *J48* foi superior. Os resultados demonstram que os algoritmos de árvores de decisão são eficientes na predição de unidades fisiográficas, apresentando valores de índice Kappa considerados bons ou satisfatórios.

Entre as unidades fisiográficas preditas pelos algoritmos, a unidade Cume Erosional teve a melhor classificação dos pixels pelo modelo de AD em função da menor extensão e características ambientais que favoreceram o processo de predição, como por exemplo, o relevo mais plano e suave ondulado.

Com o uso de algoritmos de árvores de decisão, as variáveis ambientais mais representativas para explicar a ocorrência das unidades fisiográficas em uma região com predomínio de fases de relevo ondulado e forte ondulado foram a distância diagonal da drenagem, área de contribuição, altitude e declividade.

4. ESTUDO 2 – INDIVIDUALIZAÇÃO DE CLASSES DE SOLOS POR DESAGREGAÇÃO DE POLÍGONOS DE MAPA FISIOGRAFICO.

4.1. Introdução

Nos levantamentos de solos, o pedólogo constrói mapas conceituais a partir de seu conhecimento acerca dos fatores de formação do solo, aliados à fotointerpretação e observações de campo para delinear tipos de solos como polígonos em um mapa (Hudson, 1992; Hash, 2008). Assim, os levantamentos de solos são reconhecidos como a fonte de informações mais comum sobre a distribuição dos solos na paisagem (Bui, 2004). Apesar do valor intrínseco e da relevância dos dados, eles muitas vezes não atendem às demandas atuais de escala e nível de detalhamento adequado para a modelagem e aplicações ambientais (Carré et al., 2007; Ashtekar & Owens, 2013).

Entre os fatores que restringem o uso dos levantamentos de solos está a grande heterogeneidade nos métodos e sistemas de classificação, além da escala de apresentação ou nível de detalhamento não suficientes dos mapas convencionais (Indorante et al., 1996; Basher, 1997). Mesmo com essas restrições, eles ainda constituem o melhor registro disponível sobre a ocorrência e distribuição espacial dos solos, adquirindo importância como fonte de informações no treinamento dos modelos de predição de classes de solos, sendo um grande aliado ao mapeamento digital de solos (Omuto et al., 2013).

De forma geral, um levantamento de solos inclui unidades de mapeamento (UM) simples e combinadas representadas em polígonos, sendo que esta última é composta geralmente pelo arranjo de duas ou mais classes taxonômicas, além de inclusões menores de solos. O delineamento das UMs é feito a partir da interpretação visual de pares estereoscópicos, desenvolvendo um modelo conceitual de solo-paisagem para inferir as variações espaciais do solo (Dalmolin et al. 2004).

Um dos desafios enfrentados pelos pedólogos é como extrair o máximo de informações desses dados. Uma alternativa é a desagregação espacial de polígonos de mapas convencionais de solos, de modo que melhor

represente a forma como os solos verdadeiramente ocorrem no campo, individualizando e localizando as classes de solos na paisagem (Odgers et al., 2014a).

Nesse contexto, o foco principal da desagregação de polígonos de mapas convencionais de solos é gerar um mapa de solos que espacialmente possa distinguir os tipos de solos em um maior nível de detalhe. A desagregação foi identificada como uma abordagem conceitual para traduzir os dados atuais em formatos compatíveis com as necessidades modernas e com os conceitos pedológicos de formação do solo (Bui e Moran, 2001). A desagregação espacial de mapas de solos convencionais tem sido demonstrada como uma tentativa para atualizar universalmente mapas de solos (Smith et al, 2012; Wei et al, 2010) e para criar distinções de classes dentro dos limites das UMs.

O uso de árvores de decisão (AD) tem se tornado uma técnica útil na desagregação de mapas convencionais de solos (Wei et al., 2010; Haring et al., 2012; Li et al., 2012). As ADs são estruturas hierárquicas e sequenciais que agrupam os dados por meio de uma série de subdivisões binárias. Cada nó interno indica um teste de uma variável, cada ramo representa um resultado desse teste, e cada folha representa uma classe de solo. Para evitar a ramificação dos dados, é necessário realizar a poda, que é um processo que busca diminuir o volume de dados reagrupando recursivamente folhas de menor tamanho em outros subconjuntos homogêneos (Crivelenti et al., 2009; Giasson et al., 2011).

O objetivo deste estudo foi desagregar os polígonos de unidades fisiográficas combinadas, de modo a individualizar as classes de solos ocorrentes em cada unidade fisiográfica, para melhor representá-las como unidades simples e gerar mapas de solos com maior detalhe cartográfico, ampliando a aplicabilidade desses dados em demandas futuras.

4.2. Materiais e métodos

4.2.1. Área de estudo

A área de estudo foi a microbacia Córrego Tarumãzinho (Figura 11), situada na porção oeste do Estado de Santa Catarina (SC), localizando-se a 5 km da sede do município de Águas Frias, cujo acesso se faz pela rodovia SC-479. Ocupa uma área de aproximadamente 27,7 km² e apresenta cinco fases

principais de relevo: plano, suave ondulado, ondulado, forte ondulado e montanhoso (Embrapa, 2004).

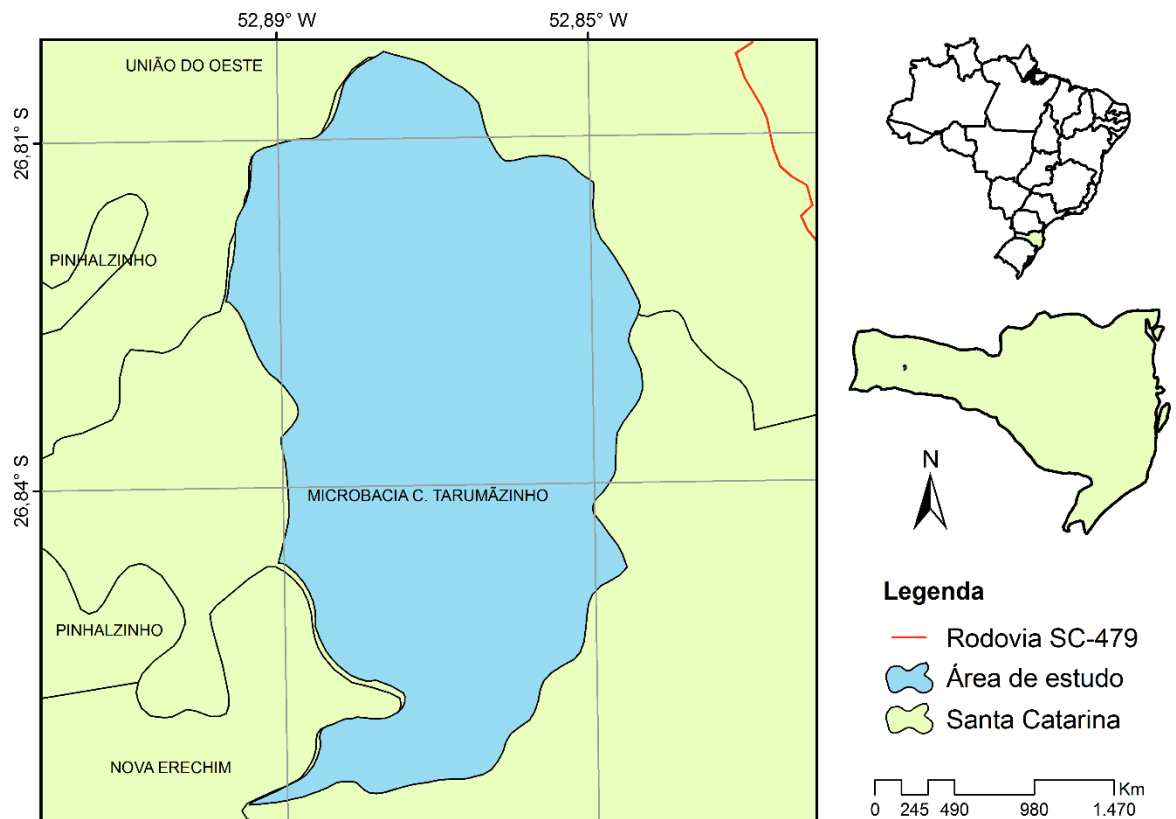


Figura 11. Localização da microbacia Córrego Tarumãzinho no oeste do Estado de Santa Catarina. Fonte: Epagri (2004).

O clima na região é classificado segundo Koppen como Cfa, caracterizado por ser constantemente úmido, com precipitação pluviométrica média anual de 1.725 mm e a temperatura média anual varia de 17,9 a 19,8°C. A temperatura média das máximas varia de 25,8 a 27,5°C, e das mínimas de 12,9 a 14,0 °C. A altitude varia de 283 metros a 660 metros. A microbacia dispõe de um mapa fisiográfico na escala de 1:25.000 e um relatório técnico, onde estão descritos os perfis dos solos dominantes nas unidades fisiográficas e suas associações (Epagri, 2004).

Os solos que compõem a microbacia tem como principal material de origem as rochas basálticas do Grupo São Bento, da Formação Serra Geral. De acordo com Silva & Bortoluzzi (1987), o material de origem constitui-se por rochas vulcânicas em derrames basálticos de textura afanítica (rochas de granulação fina, arrefecimento rápido do magma à superfície e a cristalização pode não ser completa), amigdaloidal (textura de rocha vulcânica a sub-

vulcânica contendo muitas amígdalas) no topo dos derrames, coloração cinza escura à negra, com intercalação de arenitos intertrapeanos.

Essas rochas basálticas são essencialmente constituídas por plagioclásios cálcicos e piroxênios, minerais com estabilidade bastante baixa, por este motivo alteram-se quase que totalmente a minerais argilosos com liberação de grande quantidade de óxidos, especialmente de ferro, dando origem a solos argilosos (Santa Catarina, 1986).

Uma característica marcante do basalto é a relativa facilidade de intemperização que, dependendo de outros fatores de formação do solo (principalmente relevo), dá origem a solos profundos a muito profundos. A parte superior do derrame (zona vesicular ou amigdalóide) é particularmente importante para formação dos solos. Quando essas cavidades são preenchidas com calcita, formam-se solos ricos em cálcio (Epagri, 2004).

A vegetação original direta ou indiretamente, desempenha papel importante sobre as características pedológicas da paisagem. A vegetação ocorrente na área é composta pelas seguintes formações vegetais: Floresta Estacional Decidual (Floresta Latifoliada do Alto Uruguai), localizada às margens do rio Uruguai estende-se pela microbacia até as altitudes de 500 a 600m e Floresta Ombrófila Mista, contempla o restante da microbacia nas altitudes acima de 500 metros (Santa Catarina, 1986).

4.2.2. Metodologia

O material utilizado neste estudo corresponde a um levantamento de solos e análise fisiográfica, incluindo o mapa fisiográfico na escala de 1:25.000 (Figura 12) e um relatório técnico, onde estão descritos os perfis dos solos dominantes nas unidades fisiográficas e suas associações (Epagri, 2004).

A primeira etapa do procedimento foi analisar o mapa fisiográfico e o relatório técnico, com o objetivo de identificar o número total de polígonos no mapa, o número de unidades fisiográficas (UF) que estão relacionadas a uma classe de solo e unidades fisiográficas que estão relacionadas com mais de uma classe de solo. A análise dos dados levou em consideração a posição das classes na paisagem, altitude, declividade e a descrição dos perfis georreferenciados.

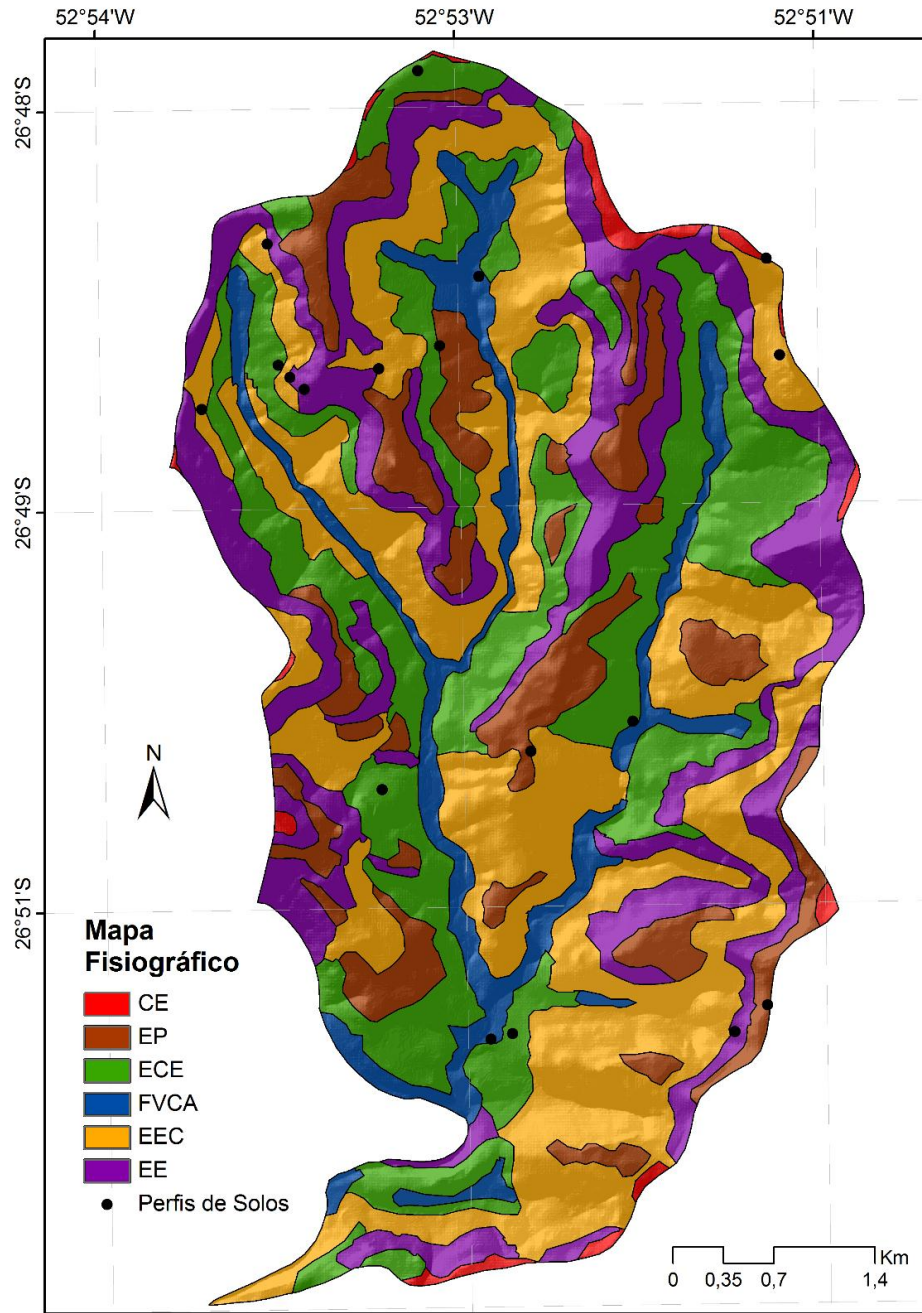


Figura 12. Mapa fisiográfico da microbacia Córrego Tarumãzinho. Fonte: Epagri (2004). Componentes: Cume Erosional (CE), Encosta Coluvial Erosional (ECE), Encosta Erosional (EE), Encosta Erosional Coluvial (EEC), Encosta em Patamar (EP), Fundo de Vale Coluvial Aluvial (FVCA).

Com base nessas informações, verificou-se que o mapa fisiográfico é constituído de 87 polígonos que representam a distribuição espacial de seis UFs, sendo quatro UFs simples e duas combinadas. Entre as UFs combinadas, uma unidade delimita duas classes de solos e a outra delimita três classes, todas na forma de associações de solos. A área ocupada pelas UFs que estão relacionadas a mais de uma classe de solo corresponde a 25,7% da área total

da microbacia (Tabela 6). No relatório técnico os solos são classificados até o quarto nível categórico, mas para este estudo as classes foram individualizadas até o segundo nível categórico por entender que a diferenciação de determinadas classes se torna difícil em função das características semelhantes e a posição que se encontram na paisagem.

Tabela 6. Descrição das unidades fisiográficas e os solos ocorrentes na microbacia Córrego Tarumãzinho. Fonte: Epagri (2004).

Unidades Fisiográficas	Solos correspondentes	Inclusões	Área	
			(%)	(ha)
Ce	Nitossolo Háplico + Nitossolo Vermelho	Cambissolo	1,4	38
Ece	Cambissolo Háplico + Nitossolo Háplico + Nitossolo Vermelho	-	24,3	669
Ee	Neossolo Litólico	Cambissolo	21,1	580
Eec	Cambissolo Háplico	Nitossolos	32,7	900
Ep	Cambissolo Háplico	-	12,3	339
FVca	Cambissolo Háplico	Gleissolos	8,2	226
Total			100,0	2752

Na etapa seguinte, foi necessário realizar um pré-tratamento dos dados antes de derivar as variáveis. No MDE, a necessidade do pré-tratamento vem do fato que diminutas imperfeições do relevo poderiam ser consideradas como ponto de sela, o que não possui qualquer sentido prático. Para contornar este problema, com a função *Focal Statistics* no ArcGIS 9.3[®], utilizou-se um filtro de médias em janela móvel retangular de 10x10 células de dimensão. Em seguida, foram removidas as depressões espúrias com o comando *Fill*.

Foram derivados a partir do MDT com resolução espacial de 5 m, um mapa com classes de declividades, para representar as fases de relevo (IBGE, 2015) e um mapa com classes de geofoma (MAcMillan, 2003) para caracterizar as diferentes formas do terreno, além do próprio mapa de elevação identificando

a altitude de ocorrência de cada classe de solo. Levando em consideração que todas as informações dos mapas utilizados são correspondentes aos dados do relatório técnico referentes a cada classe de solo. Os limites dos polígonos das UFs não foram respeitados porque polígonos de bordas mais detalhadas tendem a reduzir a inclusão de solos de unidades de mapeamento adjacentes e, conseqüentemente, a incerteza.

No ArcGIS 9.3[®] utilizou-se a função *Tabulate area*, onde os mapas de declividade, elevação, geoforma e os limites de cada polígono das UFs foram submetidos a uma tabulação cruzada. A tabulação cruzada, também conhecida como tabela de contingência, descreve simultaneamente duas ou mais variáveis e reflete a distribuição conjunta das condições destas variáveis. O cruzamento de informações só é possível quando as duas variáveis estiverem na mesma unidade observacional (ou experimental). A análise cruzada buscou individualizar em cada UF as combinações existentes entre as classes de solos.

Foram criadas regras para individualizar as áreas de ocorrência típicas de cada classe de solo nas 6 UFs. O conjunto de regras únicas foram aplicadas a todas as UFs porque solos que ocorrem em associações foram citados ocorrendo em inclusões nas unidades simples de mapeamento. As regras foram traduzidas para expressões lógicas em SIG a fim de produzir, a partir dos limites das UFs, das declividades, das geoformas e da elevação, um novo mapa somente com os *pixels* representativos de cada solo individual.

Para realizar a amostragem das variáveis, criou-se 7085 pontos amostrais aleatórios estratificados sobre o mapa fisiográfico da microbacia, equivalente à densidade amostral de 2,5 pontos por hectare. Para essa amostragem, foi usado o módulo *Sample* encontrado na extensão *Spatial Analyst* em ambiente ArcGIS 9.3[®].

Após criados os pontos, as informações ambientais referentes a cada ponto amostral foram exportadas e convertidas em um arquivo delimitado por vírgula (CSV format - *Comma Separated Values*) para adequação e uso no software Weka 3.6.6 (*Waikato Environment for Knowledge Analysis*) e usadas no treinamento do algoritmo de árvore de decisão (Hall et al., 2009). O pacote Weka consiste de uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Pode ser usado para aplicar métodos de

aprendizado a um conjunto de dados e analisar a saída para extrair informações a partir dos dados de entrada (Hall et al., 2009).

O mapa de desagregação foi gerado a partir da implementação das regras de classificação proveniente dos modelos de AD no ArcGIS 9.3® utilizando a função *Raster Calculator* até chegar aos nós finais e gerar camadas com a delimitação de cada classe de solo. Para implementação da árvore em ambiente ArcGIS as regras de classificação foram ajustadas a partir de um algoritmo elaborado em C# (C Sharp) de forma a serem diretamente utilizadas em programas SIG.

A concordância do mapa desagregado com o mapa fisiográfico foi avaliado com o uso de cinco métodos de validação para avaliar a eficácia da desagregação, considerando: acurácia geral (AG), que é a proporção dos pixels corretamente classificados em relação ao número total de pixels; erro absoluto; erro relativo entre a fração do mapa fisiográfico em relação ao mapa desagregado e comparação visual dos polígonos desagregados. Além destes, o mapa desagregado também foi validado com base em um conjunto de perfis georreferenciados, coletados em um estudo de levantamento e mapeamento de solos baseados na análise fisiográfica (Epagri, 2004).

Considerou-se a concordância espacial das coordenadas desses perfis com a localização de *pixels* individuais nos mapas desagregados e, adicionalmente, com um raio de busca (*buffer*) de 1 *pixel* (5 metros) e 2 *pixels* (10 metros) em torno de cada perfil. Neste caso, a predição foi contabilizada como correta quando a classe de solo predita em ao menos um dos *pixels* no raio de busca foi igual à observada no perfil de referência (Smith et al., 2012; Nauman & Thompson, 2014).

4.3. Resultados e discussão

Com a desagregação dos polígonos do mapa fisiográfico foi possível identificar os limites das classes de solos em cada unidade fisiográfica. As combinações mais representativas de declividade, elevação e geofoma foram selecionadas para individualizar os solos (Tabela 7) e produzir um novo mapa no formato *raster* com a localização das classes de solos na sua respectiva posição na paisagem.

Tabela 7. Regras de individualização das classes de solos nos polígonos de unidades fisiográficas combinadas, a partir de mapas de classes de declividades, elevação, geoforma e os limites dos polígonos do mapa fisiográfico.

UF	Declividade %	Elevação (m)	Geoformas	Classes de solos
CE	5	655	DSH	Nitossolo Háplico Eutrófico típico
	3	603	FSL	Nitossolo Vermelho Distrófico típico
	0-3	509	DSH	Nitossolo Vermelho Eutrófico típico
ECE	12	480	CBS	Cambissolo Háplico Alumínico típico
	9	320	BSL	Cambissolo Háplico Ta Eutrófico típico
	7	387	DBS	Nitossolo Háplico Eutrófico típico
	9	380	DBS	Nitossolo Vermelho Eutrófico latossolico
EE	27	516	DBS	Neossolo Litólico Distrófico típico
	46	496	DSH	Neossolo Litólico Eutrófico típico
EEC	21	437	CBS	Cambissolo Háplico Eutrófico típico
	30	557	DSH	Cambissolo Háplico Ta Eutrófico típico
	23	483	DBS	Cambissolo Háplico Eutrófico típico
EP	6	575	DSH	Cambissolo Háplico Ta Eutrófico típico
	11	530	CBS	Cambissolo Háplico Ta Eutrófico típico
	6	478	DSH	Cambissolo Háplico Ta Eutrófico típico
FVCA	0-3	300	BSL	Cambissolo Háplico Ta Distrófico léptico
	0-4	450	LSM	Cambissolo Háplico Ta Eutrófico típico

Componentes: DSH: divergent shoulder; BSL: back slope; DBS: divergent back slope; CBS: convergent back slope; FSL: foot slope; LSM: lower slope mound.

O mapa fisiográfico e o mapa desagregado das classes taxonômicas apresentam boa correspondência, conforme pode ser observado por comparação visual (Figura 12 e 13). Por ser uma área de pequena extensão, apenas duas UFs eram predominantemente combinadas, com diferentes arranjos de classes de solos, as demais UFs não necessitaram ser desagregadas por possuírem UFs simples. No entanto, as classes de solos ocorrentes nessas UFs foram representadas no mapa desagregado com a finalidade de obter um mapa de solos com todas as classes de solos individualizadas.

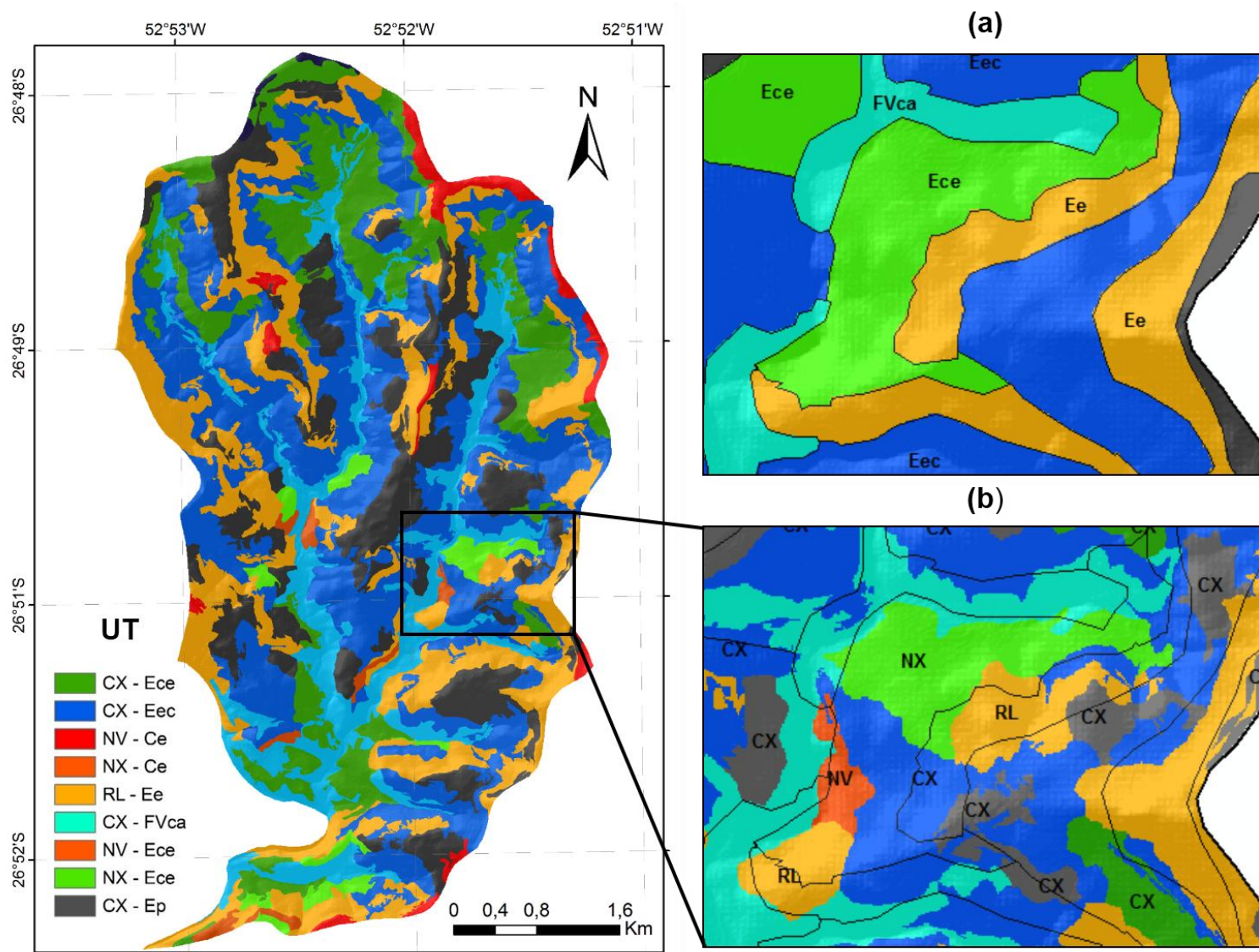


Figura 13. Mapa de unidades taxonômicas. (a) Detalhe de uma porção das UFs combinadas no mapa fisiográfico; (b) Detalhe de uma porção do mapa fisiográfico desagregado.

Considerando as UFs desagregadas, verificou-se que o erro relativo médio, ou seja, a média de quanto a fração desagregada diferiu em relação a fração do mapa fisiográfico, foi de 0,09 (Tabela 8). Isso sugere que os pontos amostrais extraídos dos polígonos das UFs combinadas foram eficientes na classificação dos *pixels* correspondentes as classes de solos, e assim puderam capturar a variabilidade e as características dessas classes. Utilizando uma técnica semelhante, Li et al. (2012) obtiveram erro relativo entre 0,49 e 0,67 para a relação entre as áreas totais de classes desagregadas e as informadas no relatório do *Canada Land Inventory* (CLI). No entanto, naquele estudo foram consideradas apenas as áreas das classes dominantes em cada UM, enquanto no presente estudo foram considerados os acertos em relação a todas as classes de solos.

Tabela 8. Valores de erro absoluto e erro relativo entre as áreas do mapa fisiográfico e a soma das áreas de classes de solos ocorrentes nas UFs, representadas no mapa de desagregação.

Classe de solo (UF)	Área do mapa fisiográfico	Área do mapa desagregado	Erro absoluto	Erro relativo
	(ha)			
NV+NX (CE)	38,0	28,7	9,30	0,24
CX+NV+NX (ECE)	669,0	670,9	1,90	0,00
(RL) EE	580,0	611,1	31,10	0,05
(CX) EEC	900,0	927,9	27,90	0,03
(CX) EP	339,0	311,2	27,80	0,08
(CX) FVca	226,0	201,8	24,20	0,11
Média			20,37	0,09

Os três casos com desvios pequenos ocorrem nas unidades fisiográficas EEC, EE e ECE (Tabela 8). Embora a unidade EEC delimite uma única classe de solo, ela ocupa cerca de 32,7% da área total da microbacia e é composta pela classe Cambissolo Háplico. A unidade EE ocupa cerca de 21,1% da área e é composta pela classe Neossolo Litólico, segundo o relatório técnico (Epagri, 2004). Já a unidade fisiográfica ECE, apresenta no mapa fisiográfico

área total de 669,0 ha, porém, ao ser desagregada a soma das classes de solos CX + NV + NX foi de 670,9 ha, ou seja, um erro absoluto de 1,9 ha, sendo que nesse caso a área da UF foi superestimada.

Ao analisar a Tabela 8, verifica-se que entre as classes desagregadas, o maior desvio ocorreu na soma das classes de solos ocorrentes na unidade CE. Essa UF apresenta no mapa fisiográfico área total de 38,0 ha e ao ser desagregada a soma das classes de solos NV + NX foi de 28,7 ha. Nas unidades desagregadas CE e ECE, o desvio pode estar associado as classes de solos ocorrentes nessas UFs, pois, tanto a unidade fisiográfica CE, quanto a unidade ECE, são compostas pelas classes Nitossolo Háplico e Nitossolo Vermelho. Logo, ao realizar a desagregação do mapa fisiográfico houve confundimento na classificação dos *pixels*. Esse erro pode ser justificável pelas características ambientais semelhantes das classes de solos que são difíceis de diferenciar.

Os erros da fração desagregada em relação a fração descrita no relatório são menores nas classes de maior extensão. Isto indica que a distribuição dessas classes de solos é consistente com as informações do relatório e representa um bom resultado, pois os desvios mais acentuados são correspondentes as menores áreas, portanto diminui as chances de erros na aplicação do mapa desagregado. Nauman e Thompson (2014), em estudo com desagregação, mostraram que o sistema de amostragem pode alterar a precisão de classes de menor extensão, se houver uma grande disparidade entre o tamanho das classes de menor e maior extensão.

Em relação a área ocupada pelas classes de solos após a desagregação (Figura 14), observa-se que os solos dominantes na microbacia são os Cambissolos Háplicos, com área total de 1895,47 ha, que ocorre principalmente nas unidades EEC e EP, seguidos dos Neossolos Litólicos, área total de 611,06 ha, que ocorre na unidade EE, em relevo bem acentuado com declividade entre 27 a 46% (Epagri, 2004).

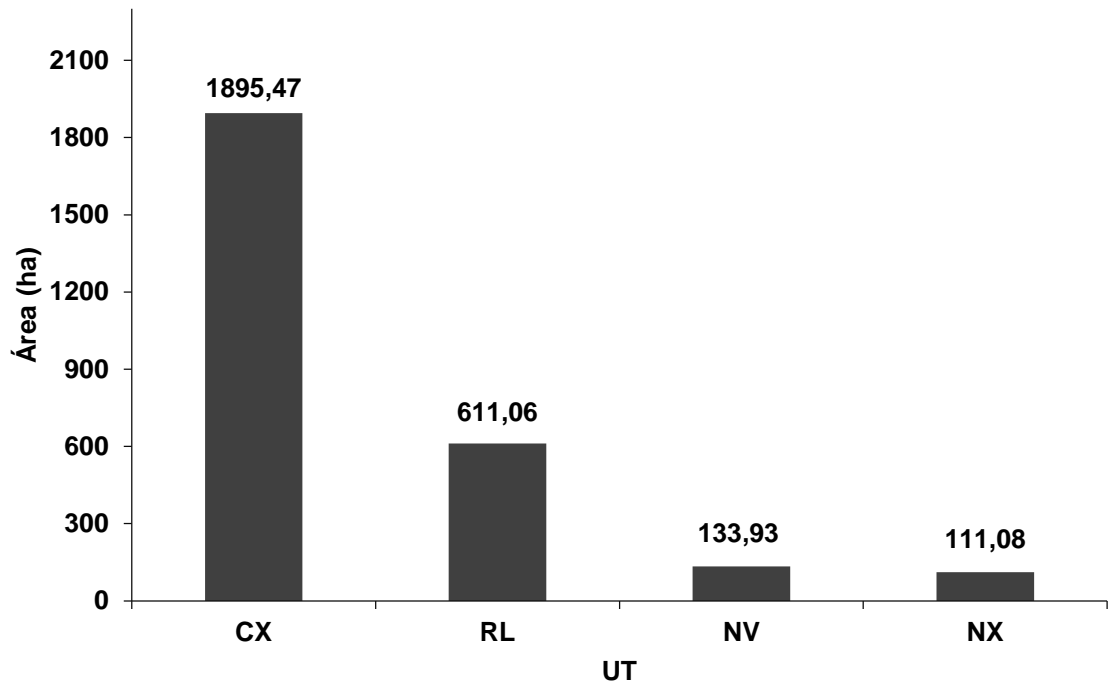


Figura 14. Distribuição da área em relação as classes de solos representadas em unidades de mapeamento simples no mapa desagregado da microbacia Córrego Tarumãzinho.

Com base nos resultados de desagregação, observa-se que apesar das áreas superestimadas em algumas UFs e subestimadas em outras UFs, a área total do mapa fisiográfico praticamente se equivale a soma das classes de solos individuais correspondentes as UFs, com R^2 de 0,99 (Figura 15). Li et al. (2012) obtiveram R^2 entre 0,87 e 0,99 no mapa desagregado em relação ao mapa convencional. Os resultados confirmam a hipótese de que a integração de variáveis ambientais preditoras com dados auxiliares pode ser útil para geração de mapas de solos mais detalhados.

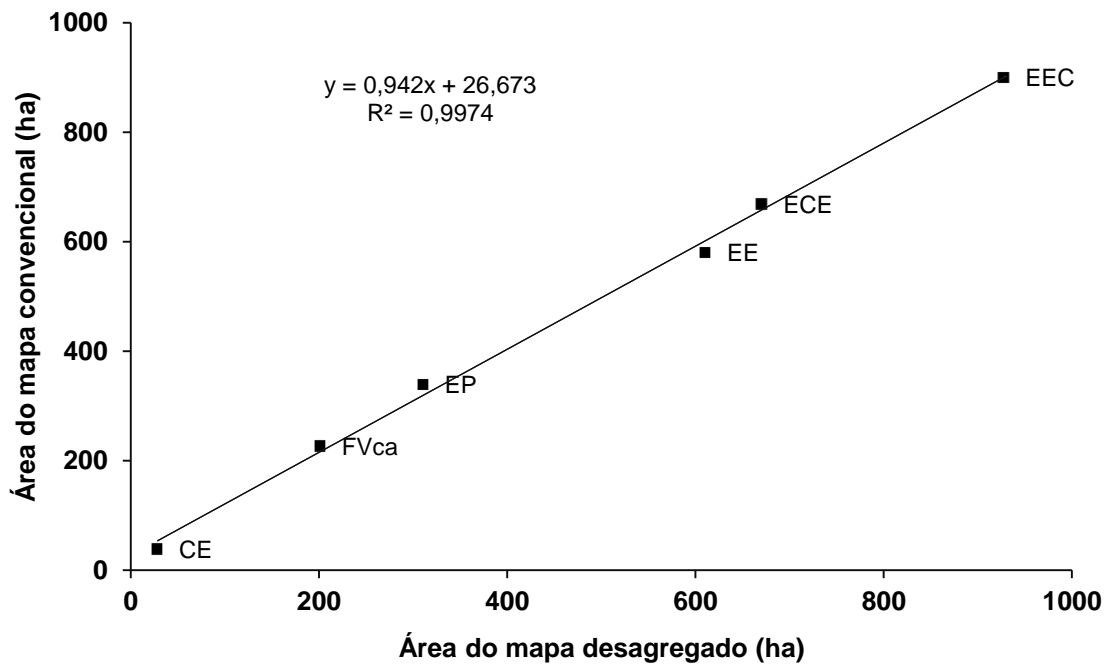


Figura 15. Relação entre a área do mapa fisiográfico e a área do mapa desagregado com as classes de solos ocorrentes representas de forma individual.

Duas unidades fisiográficas combinadas fornecem uma ilustração da abordagem da desagregação. A primeira é a unidade ECE (Tabela 9) que no mapa fisiográfico ocupa cerca de 24,3% da área total da microbacia e é composta pela associação Cambissolo Háplico, Nitossolo Háplico e Nitossolo Vermelho, segundo o relatório técnico (Epagri, 2004). Nos resultados do mapa de desagregação (Tabela 9), observa-se que a classe dominante é o Cambissolo Háplico que ocupa cerca de 67,75% da área correspondente a unidade ECE, seguido pela classe Nitossolo Vermelho 17% da área e Nitossolo Háplico 15,25% da área total.

Os Cambissolos se localizam nas áreas mais próximas ao fundo dos vales e quando localizados muito próximos às várzeas, apresentam problemas de drenagem em profundidade (abaixo de 1m), porém, não interferindo no uso com culturas anuais. Já os Nitossolos ocupam as áreas mais próximas ao final das pendentes das encostas, são profundos e bem drenados, exibindo cores vermelhas escuras nos horizontes subsuperficiais (Epagri, 2004).

A segunda UF combinada é a unidade CE que ocupa no mapa fisiográfico cerca de 1,4% da área total da microbacia e é composta pela associação Nitossolo Vermelho e Nitossolo Háptico, segundo o relatório técnico (Epagri, 2004). Na Tabela 9, observa-se que no mapa desagregado a classe dominante é o Nitossolo Vermelho que ocupa cerca de 69,36% da superfície correspondente a unidade CE, seguido pela classe Nitossolo Háptico com 30,64% da área total. Os Nitossolos estão localizados principalmente em áreas de maiores altitudes (500 a 655 metros) em relevo suave ondulado localizando-se nos divisores d'água da microbacia, caracterizam-se por serem solos profundos e com maior possibilidade de uso intensivo com culturas anuais.

Tabela 9. Área ocupada pelo mapa fisiográfico em relação a área ocupada pelas classes de solos desagregadas ocorrentes nas unidades fisiográficas.

UF	Mapa fisiográfico (ha)	Classe	Mapa desagregado (ha)	Área / unidade (%)
NX+NV (CE)	38	Nitossolo Háptico	8,8	30,64
		Nitossolo Vermelho	19,9	69,36
NX+NV+CX (ECE)	669	Nitossolo Háptico	102,3	15,25
		Nitossolo Vermelho	114,1	17,00
		Cambissolo Háptico	454,5	67,75
CX (EP)	339	Cambissolo Háptico	311,2	100,0
CX (EEC)	900	Cambissolo Háptico	927,9	100,0
NL (EE)	580	Neossolo Litólico	611,1	100,0
CX (Fvca)	226	Cambissolo Háptico	201,8	100,0
Total	2752		2751,54	

As classes de solos foram desagregadas até o segundo nível categórico do SiBCS, já que no 3º e 4º níveis do SiBCS os solos são classificados de acordo com características de mais difícil diferenciação (Tabela 9). Segundo ten Caten et al. (2012), diversos estudos de MDS buscaram mapear até o 3º e 4º níveis. No entanto, nesses níveis os acertos na distinção e na espacialização

dessas classes na paisagem podem estar mais sendo fruto do acaso do que do verdadeiro poder preditivo dos modelos.

Os resultados de acurácia geral (AG) da desagregação do mapa fisiográfico estão listados na Tabela 10. A classificação usando *pixels* dos polígonos das UFs combinadas como amostras de treinamento, atingiu AG média de 52,8% com raio de busca (*buffer*) de 5 m. Quando se considerou acertos para um raio de busca de 10 m em torno de cada perfil, a AG aumentou consideravelmente, chegando a média de 72,2% na desagregação. Esses resultados indicam que a AG aumentaria ainda mais com raios de busca maiores.

Nauman & Thompson (2014) avaliaram métodos de desagregação em dois levantamentos de solos utilizando árvores de decisão baseadas no índice de Gini, na concordância simples, obtiveram acurácia geral entre 56,4% e 65,1% da classificação dos *pixels* do mapa convencional em relação ao mapa desagregado, sendo considerado um bom resultado já que todas as unidades do mapa contêm de 10% a 25 % de outros solos ou inclusões que poderiam influenciar no resultado de acurácia.

Certas diferenças que ocorreram no mapa de desagregação em relação ao mapa fisiográfico não são necessariamente erros, são meramente desvios de uma estimativa em relação à outra. Trata-se de tentar extrair o máximo da informação existente, pois os valores informados nos relatórios frequentemente são a única referência disponível sobre a distribuição espacial das classes em mapas convencionais de solos (Odgers et al., 2014a).

Tabela 10. Medidas de acurácia da concordância dos perfis georreferenciados com a localização de pixels individuais da microbacia Córrego Tarumãzinho.

Formas de validação	Mapa de desagregação
	AG (%)
Raio de 1 pixel (5m)	52,8
Raio de 2 pixels (10m)	72,2

AG: Acurácia geral

4.4. Conclusões

O uso da desagregação possibilitou a individualização e a espacialização das classes de solos até o segundo nível categórico do Sistema Brasileiro de Classificação de Solos, sendo útil para a produção de mapas de solos mais detalhados.

Dentre as classes de solos que ocorrem na microbacia, a dominante é a dos Cambissolos Háplicos, seguidos dos Neossolos Litólicos. Os resultados demonstraram que com a desagregação de polígonos do mapa fisiográfico, foi possível prever a localização de cada classe de solo na paisagem, geração mapas com maior detalhamento, podendo ser utilizados em demandas futuras.

A desagregação foi bem sucedida nas unidades fisiográficas Encosta Coluvial Erosional, Encosta Erosional e Encosta Erosional Coluvial, apresentando os menores erros relativos em relação as demais.

5. CONSIDERAÇÕES FINAIS

Os estudos que compõem esta dissertação buscaram aplicar a técnica MDS na predição de unidades fisiográficas e desagregação de polígonos de mapa fisiográfico. O Estudo 1 destacou o uso de algoritmos de árvores de decisão e variáveis do terreno na predição de unidades fisiográficas, contribuindo para o MDS no aprimoramento de metodologias que auxiliem no mapeamento digital de unidades fisiográficas em microbacias. Dessa forma, esses resultados criam a possibilidade de aplicar a predição para áreas não mapeadas com características topográficas semelhantes.

O Estudo 1, também destacou as particularidades dos algoritmos de árvores de decisão, sendo que cada algoritmo tem uma determinada vantagem na predição de classes de solos, seja no tamanho da árvore, nos *pixels* classificados corretamente ou nos resultados de acurácia. O uso de árvores de decisão permitiu a identificação das relações entre as variáveis ambientais preditoras e a distribuição das diferentes unidades fisiográficas na microbacia.

O Estudo 2, destacou a importância dos mapas convencionais de solos no atual cenário do MDS e aplicou metodologia para obter informações mais detalhadas da distribuição dos solos na paisagem. A abordagem de

desagregação de polígonos do mapa fisiográfico se fundamentou na compreensão das informações contidas no relatório técnico no sentido de individualizar as classes de solos ocorrentes em cada unidade fisiográfica.

Apesar da área de estudo ser de 2752 ha, os resultados criam a possibilidade de melhor entendimento dos processos utilizados na desagregação e auxiliam na aplicação da técnica para áreas de maior extensão, minimizando os trabalhos adicionais de campo e representando uma econômica considerável de recursos.

6. REFERÊNCIAS

- ABDEL-KADER, F.H. Digital soil mapping at pilot sites in the northwest coast of Egypt - A multinomial logistic regression approach. **The Egyptian Journal of Remote Sensing and Space Sciences**, [El Gedida], v.14, p.29–40, 2011.
- ADHIKARI, K. et al. Constructing a soil class map of Denmark based on the FAO legend using digital techniques. **Geoderma**, Amsterdam, v. 214-215, p.101-113, 2014.
- ASHTEKAR, J.M.; OWENS, P.R. Remembering Knowledge: An Expert Knowledge Based Approach to Digital Soil Mapping. **Soil Horizons**, Madison, v.54, n.5, sept.-oct., 2013. 6p.
- BASGALUPP, M.P. **LEGAL-Tree**: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão. Tese (Doutorado) - ICMC-USP, Universidade de São Paulo, São Carlos, 2010.
- BASHER, R. Is pedology dead and buried? **Australian Journal of Soil Research**, Melbourne, v.35, p.979-994, 1997.
- BEHRENS, T. et al. Multi-scale digital terrain analysis and feature selection for digital soil mapping. **Geoderma**, Amsterdam, v.155, p.175-185, 2010.
- BEVEN, K.; KIRKBY, N. A physically based variable contributing area model of basin hydrology. **Hydrological Sciences = Bulletin des Sciences Hydrologiques**, Oxford, v.24, p.43-69, 1979.
- BISPO, P.C.; VALERIANO, M.M.; KUPLICH, T.M. Variáveis geomorfométricas locais e sua relação com a vegetação da região do interflúvio Madeira-Purus (AM-RO). **Acta Amazônica**, Manaus, v.39, n. 01, p.81-90, 2009.
- BOTERO, P.J. **Guia para el análisis fisiográfica**. Bogotá: CIAF, 1977. 88p.
- BREIMAN, L. et al. **Classification and regression trees**. Califórnia: Wadsworth & Brooks/Cole Advanced Books & Software, 1984. 368p.
- BRIGGS, D.J.; SHISHIRA, E.K. Soil variability in geomorphologically defined survey units in the Albudeite area of Murcia Province, Spain. In: JUNGRIUS, P.D. (Ed.). **Soils and geomorphology**. Cremlingen: Catena Verlag, 1985. 174p. (Catena Supplement 6). cap. 3, p.69-84.
- BUI, E.N.; LOUGHHEAD, A.; CORNER, R. Extracting soil-landscape rules from previous soil surveys. **Australian Journal of Soil Research**, Melbourne, v.37, n.3, p.495-508, 1999.
- BUI, E.N.; MORAN, C.J. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Australia. **Geoderma**, Amsterdam, v.103, p.79-94, 2001.

BUI, E.N.; MORAN, C.J. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. **Geoderma**, Amsterdam, v.111, p.21-44, 2003.

BUI, E.N. Soil survey as a knowledge system. **Geoderma**, Amsterdam, v.120, p.17-26, 2004.

CAMPOS, M.C.C. et al. Relações solo-paisagem em uma topossequência sobre substrato granítico em Santo Antônio do Matupi, Manicoré (AM). **Revista Brasileira de Ciência do Solo**, Viçosa, v.35, n.1, p.13-23, 2011.

CARVALHO, C.C.N.; FRANCA-ROCHA, W.; UCHA, J.M. Mapa digital de solos: uma proposta metodológica usando inferência fuzzy. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v.13, n.1, p.46-55, 2009.

CARVALHO JUNIOR, W. et al. Digital soilscape mapping of tropical hillslope areas by neural networks. **Scientia Agrícola**, Piracicaba, v.68, p.691-696, 2011.

CARRÉ, F. et al. Digital soil assessments: Beyond DSM. **Geoderma**, Amsterdam, v.142, p.69-9, 2007.

CHAGAS, C. DA S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. 2006. 223f. Tese (doutorado) – Universidade Federal de Viçosa, Viçosa, 2006.

CHAGAS, C. DA S. et al. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. **Pesquisa Agropecuária Brasileira**, Brasília, v.45, p.497-507, 2010.

CHRISTOFOLETTI, A. **Geomorfologia**. 2. ed. São Paulo: Edgar Blüchler, 1980.

COELHO, F.F.; GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, Santa Maria, v.40, p.1-8, 2010.

COGO, N.P. et al. Perdas de solo e água por erosão hídrica influenciadas por métodos de preparo, classes de declive e níveis de fertilidade do solo. **Revista Brasileira de Ciência do Solo**, Campinas, v.27, p.743-753, 2003.

CONGALTON, R.G. A review of assessing the accuracy of classification of remotely sensed data. **Remote Sensing of Environment**, New York, v.37, p.35-46, 1991.

CRIVELANTI, R.C. et al. Mineração de dados para a inferência de relações solo-paisagem em mapeamentos digitais de solo. **Revista Agropecuária Brasileira**, Brasília, v.44, p.1707-1715, 2009.

DALMOLIN, R.S.D. et al. Relação entre as características e o uso das informações de levantamentos de solos de diferentes escalas. **Ciência Rural**, Santa Maria, v.34, p.1479-1486, 2004.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. Rio de Janeiro, 1999. 412p.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Solos do Estado de Santa Catarina**. Rio de Janeiro: Embrapa Solos, 2004. 1 CD-ROM.; mapa color. (Embrapa Solos. Boletim de Pesquisa e Desenvolvimento; n. 46). (Anexo mapa Levantamento de reconhecimento dos Solos do Estado de Santa Catarina. Escala: 1:250.000. CDD (21. Ed.) 631.478164).

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 3. ed. Brasília, 2013. 353p.

EPAGRI-Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina S.A. **Inventário de Terras** – Microbacia Córrego Tarumãzinho, município de Águas Frias – SC, 2004. [Florianópolis, 2004].

ESRI -Environmental Systems Research Institute. **ArcGIS, Professional GIS for the desktop**, versão 9.3.1 CA. [Software]. 2009.

FAO; ITPS. **Status of the World's Soil Resources (SWSR)**. Main Report. Rome, Italy: Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, 2015.

FERREIRA, H.M. **Uso de ferramentas de aprendizado de máquina para prospecção de perdas comerciais em distribuição de energia elétrica**. 2008. Dissertação (Mestrado) - FEEC/Unicamp, 2008.

FILHO, A.C.; CURTI, N.; SHINZATO, E. Relações solo-paisagem no Quadrilátero Ferrífero em Minas Gerais. **Pesquisa Agropecuária Brasileira**, Brasília, v.45, n.8, p.903-916, ago. 2010.

FLORINSKY, I.V. et al. Prediction of soil properties by digital terrain modelling. **Environmental Modelling and Software**, Oxford, v.17, p. 295-311, 2002.

FONSECA, R.; SILVA, P.; SILVA, R. Acordo inter-juízes: O caso do coeficiente Kappa. **Laboratório de Psicologia**, Portugal, v.5, n.1, p.81-90, 2007.

FRITZSONS, E.; MANTOVANI, L.E.; AGUIAR, A.V. Relação entre altitude e temperatura: Uma contribuição ao zoneamento climático no estado do Paraná. **Revista de Estudos Ambientais**, Blumenau, v.10, n. 01, p.49-64, 2008.

GERRARD, A.J. **Soils and landforms**. London: George Allen & Unwin, 1981. 219p.

GIASSON, E. et al. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, Piracicaba, v.63, p.262-268, 2006.

GIASSON, E. et al. Decision trees for digital soil mapping on subtropical basaltic steppes. **Scientia Agricola**, Piracicaba, v.68, p.167-174, 2011.

GIASSON, E. et al. Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na bacia do Lageado Grande, RS, Brasil. **Ciência Rural**, Santa Maria, v.46, p.1967-1973, 2013.

GONZALEZ, J.P. et al. Digital Soil Mapping of Soil Properties in Honduras Using Readily Available Biophysical Datasets and Gaussian Processes. In: HARTEMINK, A.E.; MC BRATNEY, A.; MENDONÇA-SANTOS, M.L. (Org.). **Digital soil mapping with limited data**. Amsterdam: Elsevier, 2008. p.367- 380.

GOOSEN, D. **Physiography and soils of the Llanos Orientales, Colombia**. Enschede: Publications of the International Institute for Aerial Survey and Earth Sciences (ITC), 1971. 199p.

HALL, M. et al. The WEKA Data Mining Software: An Update. **SIGKDD Explorations Newsletter**, [S.l.], v.11, p.10-18, 2009.

HÄRING, T. et al. Spatial disaggregation of complex soil map units: A decisiontree based approach in Bavarian forest soils. **Geoderma**, Amsterdam, v.185-186, p.37-47, 2012.

HARTEMINK, A.E.; McBRATNEY, A. A soil science renaissance. **Geoderma**, Amsterdam, v.148, p.123-129, 2008.

HASH, S.J. **Use of Decision Tree Analysis for Predictive Soils Mapping and Implementation on the Malheur County, Oregon Initial Soil Survey**. 2008. 157 p. Dissertation (Master of Science in Soil Science) - Oregon State University, Oregon, 2008.

HENGL, T.; ROSSITER, D.G.; STEIN, A. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. **Australian Journal of Soil Research**, v.41, p.1403-1422, 2003.

HERMUCHE, P.M. et al. **Morfometria como suporte para elaboração de mapas pedológicos: I. Bacias hidrográficas assimétricas**. Planaltina: Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA, 2002. 25p. (Documentos, 68).

HÖFIG, P.; GIASSON, E.; VENDRAME, P.R.S. Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes. **Pesquisa Agropecuária Brasileira**, Brasília, v.49, p.958-966, 2014.

HUDSON, B.D. The Soil Survey as Paradigm-based Science. **Soil Science Society of American Journal**, Madison, v.56, p.836-841, 1992.

IBANEZ, J.J., ZINCK, J.A.; JIMENEZ-BALLESTA, R. Soil Science Survey: old and new challenges. In: ZINCK, J.A. **Soil survey: perspectives and strategies for 21st century**. Enschede, The Netherlands: ITC, 1993. p.7-14 (ITC Publication, n.21).

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. **Manual técnico de pedologia**. 3. ed. Rio de Janeiro, 2015. 430 p. (Manuais técnicos em geociências, n.4).

INDORANTE, S.J. et al. Positioning soil survey for the 21st century. **Journal of Soil and Water Conservation**, Ankeny, v.51, p.21-28, 1996.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Distância vertical à drenagem mais próxima.** Disponível em: <<http://www.dpi.inpe.br/Ambdata/hand.php>>. Acesso: 11 jul. 2016.

IPPOLITI, R.G.A. et al. Análise digital de terreno: Ferramenta na identificação de pedoformas em microbacia na região de “mar de morros” (MG). **Revista Brasileira de Ciência do Solo**, Viçosa, v.29, n. 2, p.269-276, 2005.

JENNESS, J. **Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v.1.2.** Jenness: Enterprises, 2006.

JENNY, H. **Factors of Soil Formation: A System of Quantitative Pedology.** New York: McGraw-Hill. 1941. 264p.

JOHNSON, D.L.; SCHAETZL, R.J. Differing views of soil and pedogenesis by two masters: Darwin and Dokuchaev. **Geoderma**, Amsterdam, v.237-238, p.176-189, 2015.

KERRY, R. et al. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. **Geoderma**, Amsterdam, v.170, p.347–358, 2012.

LAGACHERIE, P.; McBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P.; McBRATNEY, A.B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: an introductory perspective.** Amsterdam: Elsevier, 2007. p.3-22. (Developments in Soil Science, v.31).

LANDIS, J.R.; KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, Washington, v.33, n.1, p.159-174, 1977.

LI, Z. et al. Spatially locating soil classes within complex soil polygons – Mapping soil capability for agriculture in Saskatchewan Canada. **Agriculture, Ecosystems and Environment**, Amsterdam, v.152, p. 59-67, 2012.

MACMILLAN, R.A. **LandMapR© Software Toolkit- C++ Version: users manual.** Edmonton: LandMapper Environmental Solutions Inc., 2003. 110p.

MACMILLAN, R.A. Experiences with Applied DSM: Protocol, Availability, Quality and Capacity Building. In: HARTEMINK, A.E.; McBRATNEY, A.; MENDONÇA-SANTOS, M.L. (Ed.). **Digital Soil Mapping With Limited Data.** New York: Springer, 2008. p.113-135.

MANZATTO, C.V.; JUNIOR E.F.; PERES, J.R.R. (Ed.). **Uso agrícola dos solos brasileiros.** Rio de Janeiro: Embrapa Solos, 2002. 174p.

McBRATNEY, A.; BISHOP, T. **The information content of digital (soil) maps.** The Second Approximation International Conference on Soil Resources. Their Inventory, Analysis and Interpretation for Use in the 21st Century. 1999.

McBRATNEY, A.B. et al. On digital soil mapping. **Geoderma**, Amsterdam, v.17, p.3-52, 2003.

McBRATNEY, A.B. et al. Digital Soil Mapping. In: HUANG, P.M.; LI, Y.; SUMNER, M.E. **Handbook of Soil Sciences: Properties and Processes**. Boca Raton: Taylor & Francis, 2012. p.37-43.

MENDONÇA-SANTOS, M.L.; SANTOS, H.G. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHERIE, P.; McBRATNEY, A.B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: an introductory perspective**. Amsterdam: Elsevier, 2007. p.39-54. (Developments in soil science, v.31).

MENEZES, M.D. et al. Digital soil mapping approach based on fuzzy logic and field expert knowledge. **Ciência Agrotecnologia** [online], v.37, n.4, p.287-298, 2013.

MINASNY, B.; McBRATNEY, A. Digital soil mapping: A brief history and some lessons. **Geoderma**, Amsterdam, v.264, p.301-311, 2016.

MOORE, I.D.; GRAYSON, R.B.; LADSON, A.R. Digital terrain modeling – a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, Chinchester, v.5, n.1, p.3-30, 1991.

MOORE, I. et al. Soil attribute prediction using terrain analysis. **Soil Science Society American Journal**, Madison, v.57, p.443-452, 1993.

MORAN, C.J.; BUI, E. Spatial data mining for enhanced soil map modelling. **International Journal of Geographical Information Science**, London, v.16, p.533-549, 2002.

MUÑOZ, V.A. **Análise geomorfométrica de dados SRTM aplicada ao estudo das relações solo-relevo**. 2009. 112p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2009.

NATALE, F.; CARVALHO, N.; PAULRUD, A. Defining small-scale fisheries in the EU on the basis of their operational range of activity The Swedish fleet as a case study. **Fisheries Research**, New York, v.164, p. 286–292, april 2015.

NAUMAN, T.W.; THOMPSON, J.A. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. **Geoderma**, Amsterdam, v. 213, p. 385–399, 2014.

ODGERS, N.P. et al. Disaggregating and harmonising soil map units through resampled classification trees. **Geoderma**, Amsterdam, v.214, p.91–100, 2014a.

OLIVEIRA, J.G.R. Erosão no plantio direto: perda de solo, água e nutrientes. **Boletim de Geografia**, Maringá, v.30, n.3, p.91-98, 2012.

OLLIER, C.D. Terrain classification: methods, applications and principles. In: HAILS, J.R. (Ed.). **Applied geomorphology: a perspective of the contribution of**

geomorphology to interdisciplinary studies and environmental management. Amsterdam: Elsevier Scientific Publishing Company, 1977. cap. 8, p.277-316.

OMUTO, C.; NACHTERGAELE, F.; ROJAS, R.V. **State of the art report on global and regional soil information**: Where are we? Where to go? Global Soil Partnership Technical Report. Roma: FAO, 2013. 69p. Disponível em: <<http://www.fao.org/docrep/017/i3161e/i3161e.pdf>>. Acesso em: 05 jul. 2016.

PISSARA, T.C.T.; POLITANO, W.; FERRAUDO, A.S. Avaliação de características morfométricas na relação solo-superfície da bacia hidrográfica do córrego Rico, Jaboticabal (SP). **Revista Brasileira Ciência do Solo**, Viçosa, n.28, p.297-305, 2004.

QUINLAN, J.R. **C4.5: Programs for Machine Learning**. San Mateo, CA: Morgan Kaufmann, 1993. 20p.

QUINN, P.F. et al. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. **Hydrological Processes**, Chichester, v.5, p.59-79, 1991.

ROKACH, L.; MAIMON, O.Z. **Data mining with decision trees**: theory and applications. Londres: World Scientific, 2008. 244p.

SANCHEZ, P.A. et al. Digital Soil Map of the World. **Science**, Washington, v.35, p. 681-682, 2009.

SANTA CATARINA. Gabinete de Planejamento e Coordenação Geral. Subchefia de Estatística, Geografia e Informática. **Atlas de Santa Catarina**. Rio de Janeiro: Aerofoto Cruzeiro, 1986. 173p.

SANTOS, H.G. Country Report: Brazil. In: SOIL Survey: perspectives and strategies for the 21st century: **International Workshop for heads of national soil survey organizations**. Enschede, The Netherlands, 1992. (ITC Publication n. 21) p.61-67.

SANTOS, H.G. et al. Distribuição Espacial dos Níveis de Levantamento de Solos no Brasil. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 34., 2013, Florianópolis. **Anais...** Florianópolis, 2013.

SARMENTO, E.C. et al. Prediction of soil orders with high spatial resolution: response of different classifiers to sampling density. **Pesquisa Agropecuária Brasileira**, Brasília, v.47, n.9, p.1395–1403, 2012.

SARMENTO, E.C. **Predição de classes de solos em diferentes escalas na Serra Gaúcha usando mapeamento digital de solos a partir de dados legados**. 2015. 116 f. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS, 2015.

SATO, L.Y. et al. Análise comparativa de algoritmos de árvores de decisão do sistema WEKA para classificação do uso e cobertura da terra. In: SIMPÓSIO

BRASILEIRO DE SENSORIAMENTO REMOTO - SBSR, 16., 2013, Foz do Iguaçu, PR, Brasil. **Anais...** Foz do Iguaçu, PR, Brasil : INPE, 2013.

SCULL, P.; FRANKLIN, J.; CHADWICK, O.A. The application of classification tree analysis to soil type prediction in a desert landscape. **Ecological Modelling**, Amsterdam, v.181, p.1-15, 2005.

SCULL, P.; FRANKLIN, J.; CHADWICK, O.A. Serviço Nacional de Levantamento e Conservação de Solos. **Bases para leitura de mapas de solos**. Rio de Janeiro: EMBRAPA-SNLCS, 1981. (Série Miscelânea, 4).

SHI, H. **Best-first decision tree learning**. Master (Thesis) - University of Waikato, Hamilton, NZ. COMP594, 2007.

SILVA, L.C.; BORTOLLUZI, C.A. **Texto explicativo para o Mapa Geológico do Estado de Santa Catarina** - E: 1/500.000. Florianópolis: DNPM/CPRM, 1987. 216p.

SILVA, C.C. **Mapeamento digital de classes de solo**: aplicação de metodologia na folha Botucatu (sf-22-z-b-iv-3) e validação de campo. 2012. 117 fls. Dissertação (Mestrado em Agricultura Tropical e Subtropical) – Instituto Agrônômico, Campinas, 2012.

SIRTOLI, A.E. et al. Oka-Fiori. Atributos do relevo derivados de modelo digital de elevação e suas relações com solos. **Scientia agraria**, Curitiba, v.9, p.317-329, 2008.

SMITH, C.A.S. et al. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps. In: MINASNY, B.; MALONE, B.P.; McBRATNEY, A.B. (Ed.). **Digital Soil Assessments and Beyond**. London: Taylor & Francis, 2012, p.215–220.

SUBBURAYALU, S.K.; JENHANI, I.; SLATER, B.K. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. **Geoderma**, Amsterdam, v. 213, p. 334–345, 2014.

SYSTEM FOR AUTOMATED GEOSCIENTIFIC ANALYSES. **SAGA. Version: 2.1.4**. [Software] Copyrights (c) 2002-2014 by Olaf Conrad. GNU General Public License version 2.0. 1999. Available at: <http://www.saga-gis.org>. Access em: 2 Abr. 2016.

TAN, P.N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**. Boston, MA: Addison-Wesley Longman Publishing Co., 2005.

TEN CATEN, A. et al. Mapeamento digital de solos através da aplicação de componentes principais em modelos logísticos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, Natal, RN, 2009. **Anais...** Natal, RN: INPE, 2009. p.7677-7684.

TEN CATEN, A. et al. Extrapolação das relações solo-paisagem a partir de uma área de referência, **Ciência Rural**, Santa Maria, v.41, p.812-816, 2011a.

TEN CATEN, A.; DALMOLIN, R.S.D.; RUIZ, L.F.C. Digital soil mapping: strategy for data pre-processing. **Revista Brasileira de Ciência do Solo**, Viçosa, v.36, p.1083-1091, 2012.

TEN CATEN, A. et al. Mapeamento digital de classes de solos: características da abordagem brasileira. **Ciência Rural**, Santa Maria, v.42, n.11, p.1989-1997, 2013.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparação do uso de modelos digitais de elevação em mapeamento digital de solos em Dois Irmãos, RS, Brasil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, p.1367-1376, 2014.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparação de esquemas de amostragem para treinamento de modelos preditores no mapeamento digital de classes de solos. **Revista Brasileira de Ciência do Solo**, Viçosa, v.39, p.14-20, 2015.

VALERIANO, M.M. **Topodata**: guia para utilização de dados geomorfométricos locais. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2008.

VILLOTA, H. El sistema CIAF de clasificación fisiográfica del terreno. **Revista CIAF**, v.13, n.1, p. 55-70, 1992.

VILLOTA, H. **Geomorfología aplicada a levantamientos edafológicos y zonificación física de tierras**. Bogotá: Instituto Geográfico Agustín Codazzi, 2005. 184p.

WEI, S. et al. Digital Harmonisation of Adjacent Analogue Soil Survey areas – 4 Iowa Counties. In: WORLD CONGRESS OF SOIL SCIENCE, SOIL SOLUTIONS FOR A CHANGING WORLD, 19., Brisbane, 2010. **Proceedings...** Brisbane: International Union of Soil Science, 2010. 1 DVD.

WEISS, A. **Topographic Position and Landforms Analysis**. 2001. (Poster presentation, ESRI User Conference, San Diego, CA, 2001).

WILSON, J.P.; GALLANT, J.C. (Ed.). **Terrain Analysis: principles and applications**. New York: John Wiley & Sons, 2000. 479p.

WITTEN, I.H.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2nd ed. São Francisco: Morgan Kaufmann, 2005. 560p.

WU, Q. et al. TrackerDetector: A system to detect third-party trackers through machine learning. **Computer Networks**, Amsterdam, v. 91, p.164–173, nov. 2015.

WYSOCKI, D. A. et al. Soil Landscape Models. In: HUANG, P.M.; LI, Y.; SUMNER, M.E. **Handbook of Soil Sciences: properties and processes**. Boca Raton: Taylor & Francis, 2012. p.29-4–29-8.