

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

**MAPEAMENTO DIGITAL DE CLASSES DE SOLOS: DENSIDADE DE
AMOSTRAGEM, SELEÇÃO DE VARIÁVEIS E EXTRAPOLAÇÃO PARA
ÁREAS FISIOGRAFICAMENTE SEMELHANTES**

Tatiane Bagatini
Tese de doutorado

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

**MAPEAMENTO DIGITAL DE CLASSES DE SOLOS: DENSIDADE DE
AMOSTRAGEM, SELEÇÃO DE VARIÁVEIS E EXTRAPOLAÇÃO PARA
ÁREAS FISIOGRAFICAMENTE SEMELHANTES**

Tatiane Bagatini
Engenheiro Agrônomo (UFRGS)
Mestre em Ciência do Solo (UFRGS)

Tese apresentada como
um dos requisitos à obtenção do
Grau de doutor em Ciência do Solo

Porto Alegre (RS), Brasil

29 de outubro de 2015

Bagatini, Tatiane

MAPEAMENTO DIGITAL DE CLASSES DE SOLOS: DENSIDADE DE AMOSTRAGEM, SELEÇÃO DE VARIÁVEIS E EXTRAPOLAÇÃO PARA ÁREAS FISIOGRAFICAMENTE SEMELHANTES / Tatiane Bagatini. -- 2015.

107 f.

Orientador: Elvio Giasson.

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.

1. Pedologia. 2. Classes de Solos. 3. Mapeamento Digital de Solos. 4. Árvores de decisão. 5. Acurácia geral. I. Giasson, Elvio, orient. II. Título.

MAPEAMENTO DIGITAL DE CLASSES DE SOLOS: DENSIDADE DE AMOSTRAGEM, SELEÇÃO DE VARIÁVEIS E EXTRAPOLAÇÃO PARA ÁREAS FISIOGRAFICAMENTE SEMELHANTES

Tese de Doutorado 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 Doutora em Ciência do Solo.

Aprovada em 29 de outubro de 2015

Homologada em.....

BANCA EXAMINADORA:

Prof. Paulo César do Nascimento
UFRGS

PhD. Egon Klamt
UFRGS

PhD. Ivan Luiz Zilli Bacic
EPAGRI

Orientador - Prof. Elvio Giasson
UFRGS

Ao meu esposo, Evandro Conte
e minha filha, Alice .

DEDICO

AGRADECIMENTOS

Ao professor Elvio Giasson, pela orientação, dedicação, incentivo, apoio incondicional.

Aos professores do Programa de Pós-Graduação em Ciência do Solo da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul (PPGCS/FA/UFRGS), pela transmissão de conhecimentos, disponibilidade a qualquer hora, bem como pela convivência e amizade.

Ao PPGCS e ao Departamento de Solos da FA/UFRGS, pela oportunidade para realização do curso.

Ao CNPq, pela concessão da bolsa de pós-graduação.

Aos meus pais, Mário João Bagatini e Teresinha Bagatini, pelo apoio incondicional e pelos ensinamentos.

Aos meus irmãos, Cleimar e Fernanda, pelas suas existências e importâncias na minha vida, bem como pelo apoio.

Ao meu marido e minha filha, Evandro e Alice. Ao meu marido pelo companheirismo, carinho, atenção e apoio durante toda o período do doutorado e a minha filha por tornar esse momento mais leve, porém um pouco mais trabalhoso.

Aos amigos e colegas do PPGCS/UFRGS, em especial aos colegas da gênese e classificação dos solos, Rodrigo, Benito, Pedro, Joelma, Jéssica, Nahiana, Tiago, Catiline e Paula, por repartirem expectativas, anseios, objetivos e conhecimentos e ensinamentos, bem como por compartilharem dessa oportunidade única, de convivência, de construção pessoal e profissional, levado a efeito tudo de forma muito especial e com muita ternura e muito carinho.

MAPEAMENTO DIGITAL DE CLASSES DE SOLOS: DENSIDADE DE AMOSTRAGEM, SELEÇÃO DE VARIÁVEIS E EXTRAPOLAÇÃO PARA ÁREAS FISIOGRAFICAMENTE SEMELHANTES^{1/}

Autor: Tatiane Bagatini

Orientador: Elvio Giasson

RESUMO

Nas últimas décadas o Mapeamento Digital de Solos (MDS) está ganhando espaço devido ao aumento da disponibilidade de dados numéricos, entretanto algumas metodologias ainda precisam ser definidas, dentre elas, a densidade de amostras e as variáveis a serem utilizadas para a alimentação dos modelos. Assim, o objetivo deste trabalho foi testar diferentes densidades de amostragem, diferentes conjuntos de variáveis para avaliar a resposta dos modelos e, a partir disso, realizar a extrapolação de classes de solos em paisagens semelhantes. Esta tese é composta de três estudos que testam metodologias para a predição de ocorrência de classes de solos utilizando técnicas do MDS. No primeiro estudo, realizado nas bacias do Santo Cristo e do Lageado Grande foi testado o efeito de diferentes densidades de amostragens sobre a capacidade preditiva dos modelos de predição de ocorrência de classes de solos. Os modelos preditores foram treinados com dados dos atributos do terreno derivados do modelo digital de elevação e com informações de solos extraídas do mapa pedológico. De modo geral o aumento da densidade de amostragem resultou no aumento da concordância com os mapas de referências e no aumento do número de unidades de mapeamento preditas. No segundo estudo, realizado nas bacias do Santo Cristo e do Arroio Portão, foi testado o efeito de diferentes conjuntos de variáveis geomorfométricas sobre a capacidade preditiva dos modelos de predição de ocorrência de classes de solos. A partir do modelo digital de elevação foram geradas onze variáveis. As variáveis comprimento de fluxo, elevação do terreno e distância de rios foram as que mais influenciaram os resultados de acurácia e a quantidade de unidades de mapeamento preditas. O modelo gerado com somente as três variáveis gerou o modelo com resultados semelhantes ao modelo gerado com todas as variáveis. O terceiro estudo foi realizado em duas etapas. A primeira etapa foi realizada nas bacias do Santo Cristo e do Arroio Portão e a segunda nas bacias do Santo Cristo e na do Lageado Grande. Na primeira etapa dividiu-se as bacias em partes iguais, utilizando-se uma para o treinamento e a outra para a validação dos modelos. A acurácia foi maior na área de treinamento do que na área de validação, entretanto mostrou-se uma ferramenta interessante a ser utilizada para a elaboração de mapas. Na segunda etapa utilizou-se a bacia do Santo Cristo para o treinamento do modelo e a do Lageado Grande para a validação. Nesta fase os modelos não conseguiram gerar mapas com boas acurácias na área de validação.

^{1/} Tese de doutorado em Ciência do Solo, Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS (106) – Outubro, 2015.

DIGITAL SOIL CLASS MAPPING: SAMPLING DENSITY, VARIABLE SELECTION AND EXTRAPOLATION TO PHYSIOGRAPHIC SIMILAR AREA^{1/}

Author: Tatiane Bagatini

Adviser: Elvio Giasson

SUMMARY

In recent decades the digital soil mapping is gaining ground due to the increase of numerical data, but some methodologies have yet to be defined, among them, the density of sampling and the selection of attributes models training. The objective of this study was to evaluate the use of different methodologies and materials for data analysis and prediction of occurrence of soil classes. This thesis is composed of three studies testing methodologies for soil classes of occurrence of prediction using MDS techniques. In the first study, performed in the Santo Cristo and the Lageado Grande watersheds, it was tested the effect of different densities of samples on the capacity of models to predict the occurrence of soil classes. The predictive models were trained with data attributes derived from a digital terrain elevation model and information extracted from a legacy soil map. In general, the increase in sampling density resulted in an increase in accordance with the reference map and increase the number of predicted map units. In the second study, performed in the Santo Cristo and Arroio Portão watersheds, it was tested the effect of different sets of geomorphometric variables on the predictive ability of occurrence of prediction models of soil classes. From the digital elevation model were derivate eleven variables. The variables flow length, altitude and distance from rivers were the ones that more influenced the results of accuracy and number of predicted mapping units. The model generated with only this three variables generated the model with results similar to the model that used all variables. The third study was performed in two stages. The first step performed in the Santo Cristo and Arroio Portão watersheds, and the second in the Santo Cristo and Lageado Grande watersheds. In the first step the basins were divided into two equal parts using a part for model training and the other for model validation. The accuracy was greater in the training area than in the validation area; however the extrapolation to similar areas proved to be an interesting tool to be used for the preparation of maps. In the second step it was used the basin of the Santo Cristo for model training and the Lageado Grande for model validation. In this step the models were not able to generate maps with good accuracy in the validation area.

^{1/} Doctoral thesis in Soil Science, Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS (106 p) – Outubro, 2015.

SUMÁRIO

	Página
1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA	4
2.1. Mapeamento Convencional de Solos versus Mapeamento Digital de Solos.....	4
2.2. Modelos preditivos empregados no MDS	5
2.3. Variáveis utilizadas para o treinamento dos modelos.....	7
2.3.1. Classificação e caracterização das variáveis geomorfométricas ...	8
2.3.2 Variáveis primárias.....	9
2.3.3. Variáveis secundárias	11
2.4. Avaliação da acurácia dos mapas gerados por MDS.	12
3. MATERIAL E MÉTODOS.....	21
3.1. Localização das áreas de estudo.....	21
3.1.1. Bacias do Lageado Grande e Santo Cristo	21
3.1.3. Bacia do Arroio Portão	27
3.2. Procedimentos gerais para a obtenção das variáveis geomorfométricas e aplicação dos modelos	30
4. CAPÍTULO I. SELEÇÃO DE DENSIDADE DE AMOSTRAGEM PARA CALIBRAÇÃO DE MODELOS DE ÁRVORE DE DECISÃO PARA O MAPEAMENTO DIGITAL DE CLASSES DE SOLOS.....	32
4.1 Introdução	32
4.2 Materiais e Métodos.....	33
4.3 Resultados e discussão	35
4.4 Conclusões	41
5. CAPÍTULO II – RESPOSTA DOS MODELOS DE ÁRVORE DE DECISÃO COM DIFERENTES CONJUNTOS DE VARIÁVEIS UTILIZADAS EM MAPEAMENTO DIGITAL DE SOLOS	42
5.1 Introdução	42
5.2 Materiais e métodos.....	44
5.3 Resultados e discussão	45
5.4 Conclusões	61
6. CAPÍTULO III – EXPANSÃO DE MAPEAMENTO DE SOLO PARA ÁREAS FISIOGRAFICAMENTE SEMELHANTES A PARTIR DE UM MODELO DE PREDIÇÃO TREINADO EM ÁREA DE REFERÊNCIA	62
6.1 Introdução	62
6.2 Material e métodos	64
6.3 Resultados e discussão	66
6.4 Conclusões	77
7. CONCLUSÕES GERAIS.....	78
8. REFERÊNCIAS	80
9. APÊNDICES	88

RELAÇÃO DE TABELAS

	Página
1. Relação de artigos referentes ao mapeamento digital de classes de solos	14
2. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Lageado Grande: simbologia usada no mapa de solos e a área e representatividade de ocorrência da UM.....	23
3. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Santo Cristo: simbologia usada no mapa de solos e a área e representatividade de ocorrência da UM.....	25
4. Legenda de identificação das Unidades de Mapeamento (UM) da bacia Santo Cristo e da bacia do Arroio Portão: composição taxonômica das UM, área e representatividade de ocorrência da UM.....	28
5. Resultados do tamanho das árvores de decisão, número de UMS, acurácia geral (AG) e UMS preditas na bacia do Santo Cristo dos modelos gerados com somente um das variáveis.....	48
6. Resultados do tamanho das árvores de decisão, número de UMS, acurácia geral (AG) e UMS preditas na bacia do Arroio Portão dos modelos gerados com somente um das variáveis.....	49
7. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo retirando uma das variáveis do conjunto de dados	50
8. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Arroio Portão retirando uma das variáveis do conjunto de dados.	51
9. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo com a combinação de variáveis duas a duas para geração do modelo de árvore de decisão.	54
10. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Arroio Portão combinação de variáveis duas a duas para geração do modelo de árvore de decisão.	57
11. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo e do Arroio Portão com as variáveis comprimento de fluxo (CF), altitude (MNT) e distância de rios (DIST) para geração do modelo de árvore de decisão.	60
12. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Lageado Grande: simbologia usada no mapa de solos e a sua reclassificação de equivalência com a bacia do Santo Cristo.....	66

13. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento e da área de validação da Bacia Santo Cristo.....	68
14. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento e da área de validação da Bacia do Arroio Portão.	72
15. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento (bacia Santo Cristo) e da área de validação (bacia Lageado Grande).	76

RELAÇÃO DE FIGURAS

	Página
1. Esquema da direção e acúmulo de fluxo.	10
2. Esquema das classes de declividade propostas por Jenness. Adaptado de Jenness (2006).....	12
3. Mapa de localização das bacias utilizadas no estudo.....	21
4. Mapa de solos (a) e mapa de relevo (b) da bacia do Lageado Grande com suas respectivas legendas.....	24
5. Mapa de solos (a) e mapa de relevo (b) da bacia do Santo Cristo com suas respectivas legendas.....	26
6. Mapa de solos (a) e mapa de relevo (b) da bacia do Arroio Portão com suas respectivas legendas.....	29
7. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos (UMS) preditas (b) e acurácia geral nas diferentes densidades de amostragens na Bacia do Santo Cristo.....	37
8. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos preditas (UMS) (b) e acurácia geral (c) nas diferentes densidades de amostragens na Bacia do Lageado Grande.....	38
9. Figura esquemática da localização das áreas de treinamento e de validação das Bacias do Santo Cristo e do Arroio Portão.	64
10. Mapa da bacia Santo Cristo com a distribuição original das classes de solos (a) e o mapa de solos predito (b).	70
11. Mapa da bacia Arroio Portão com a distribuição original das classes de solos (a) e o mapa de solos predito (b).	73
12. Mapa da bacia Lageado Grande com a distribuição original das classes de solos (a) e o mapa de solos predito (b).....	77

RELAÇÃO DE APÊNDICES

Página

1. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área interna da bacia do Lageado Grande, com diferentes números de elementos no nó final (M).....89
2. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área maior da bacia do Lageado Grande, com diferentes números de elementos no nó final (M).....90
3. Resultados referentes ao número de folhas (NF), tamanho das árvores de decisão (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados normalizados na área maior da bacia do Lageado Grande, com diferentes números de elementos no nó final (M).....91
4. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados normalizados na área maior da bacia do Santo Cristo, com diferentes números de elementos no nó final (M)92
5. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área interna da bacia do Santo Cristo, com diferentes números de elementos no nó final (M).....93
6. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área maior da bacia do Santo Cristo, com diferentes números de elementos no nó final (M).....94

1. INTRODUÇÃO

O solo é um recurso natural de fundamental importância, pois, além de servir de suporte a todas as espécies e edificações, constitui o principal meio para a produção de alimentos na Terra. Entretanto, devido ao aumento da exploração biótica, desmatamentos e, principalmente, o uso indevido, como preparos intensos, pastoreio intensivo, cultivo excessivo, monocultura, entre outros, está ocorrendo a degradação dos solos. Sob tais condições, a melhoria e a sustentabilidade da produção agrícola exigem um monitoramento mais criterioso, que pode ser possível através de avaliações e mapeamento dos recursos básicos, sendo um deles, o solo.

As informações sobre os solos no que diz respeito à sua distribuição espacial e suas características físico-química-biológicas é de suma importância na formulação de qualquer plano de uso do solo. Neste sentido, os levantamentos de solos ajudam a coletar informações sobre a sua gênese, extensão, potencialidades e limitações, podendo assim, prever seu comportamento e determinar para que finalidade específica o mesmo pode ser utilizado. Para tanto, necessita-se de mapas em escalas adequadas ao planejamento de pequenas áreas.

No Brasil a disponibilidade de mapas de solos em grandes escalas é muito pequena, pois na totalidade do seu território, possui apenas mapas de solos na escala 1:1.000.000, que foram gerados através de levantamentos publicados no Projeto Radambrasil (Ker & Almeida 1986). Por outro lado, o Estado do Rio Grande do Sul, como quase todos os Estados, possui mapas de solos recobrimdo todo seu território na escala 1:750.000, resultante do Levantamento de Reconhecimento de Solos do Estado do Rio Grande do Sul (BRASIL, 1973). Esses mapas, apesar de serem úteis para projetos em grandes áreas, não servem para planejamento de uso dos solos em uma propriedade rural ou bacias hidrográficas (DALMOLIN et al, 2004). Para tanto,

necessita-se de mapas de solos em escalas maiores que 1:100.000 que, para o Estado do Rio Grande do Sul, existem apenas em alguns municípios ou bacias hidrográficas.

Apesar de haver uma crescente necessidade de informação de mapas de solos de maiores escalas, os custos e o tempo necessário para a realização e execução, bem como as necessidades de recursos humanos envolvidos, tornam esses levantamentos de difícil execução. Neste sentido, devido à necessidade de diminuição de tempo e recursos financeiros para a realização dos mapas com nível de detalhamento mais adequado, encontrou-se em métodos matemáticos computacionais, associados a parâmetros derivados de modelos numéricos de terreno e, conjuntamente, com um sistema de informação geográfica, a solução para a obtenção desses mapas (Hengl e Rossiter, 2003). Entretanto, apesar dessas novas metodologias serem promissoras para a elaboração dos mapas, algumas metodologias ainda precisam ser melhor elucidadas. Dentre os aspectos metodológicos que ainda precisam ser mais estudados se encontra a determinação da melhor densidade de amostragem. Para os mapas convencionais existe uma metodologia que determina a quantidade de perfis ou amostras por unidade de área que deve ser utilizada para cada escala de mapeamento, entretanto, para o mapeamento digital não existe essa padronização. Além da falta de padronização da densidade de amostragem, a seleção das variáveis a serem utilizadas para a alimentação de modelos é outro fator em que não há um consenso sobre quais são as mais adequadas a serem utilizadas. Assim sendo, neste trabalho foram testados algumas densidades de amostragens e um conjunto de variáveis tentando buscar uma padronização para esses temas pendentes. Posteriormente realizou-se a extrapolação dos mapas de solos utilizando-se os resultados da seleção da densidade de amostragem e seleção das variáveis.

Assim sendo o presente trabalho foi estruturado em forma de estudos. O primeiro estudo apresentado no capítulo I é intitulado: “*Seleção de densidade de amostragem para calibração de modelos de árvore de decisão para o mapeamento digital de classes de solos.*” Neste estudo buscou-se definir qual a densidade de amostragem que resulta em modelos com melhor acurácia e, para tanto, utilizou-se duas bacias hidrográficas situadas no

nordeste do Rio Grande do Sul, sendo elas a bacia do Lageado Grande e a bacia do Santo Cristo.

O segundo estudo apresentado no capítulo II é intitulado de “Resposta dos modelos de árvore de decisão com diferentes conjuntos de variáveis utilizadas em mapeamento digital de solos”. Neste estudo buscou-se determinar quais as variáveis que mais contribuem para a melhor resposta dos modelos. Para tanto utilizou-se duas bacias com características distintas de relevo e solos, uma localizada no nordeste do Estado do Rio Grande do Sul, a bacia do Santo Cristo, e a outra localizada na região central do Estado, a bacia do Arroio Portão.

O terceiro estudo apresentado no capítulo III é intitulado “Expansão de mapeamento de solo para áreas fisiograficamente semelhantes a partir de um modelo de predição treinado em área de referência”, no qual testou-se a viabilidade de extrapolação de mapas de solos. Esse estudo foi realizado em duas etapas. Na primeira foram utilizadas as bacias mencionadas no capítulo II, as quais foram divididas em duas partes. Uma parte da bacia serviu para a calibração do modelo e outra para a validação do modelo. Na segunda etapa foram utilizadas as bacias mencionadas no capítulo I, nas quais, uma das bacias foi utilizada para a calibração e a outra para a validação do modelo.

2. REVISÃO BIBLIOGRÁFICA

2.1. Mapeamento Convencional de Solos versus Mapeamento Digital de Solos

De acordo com Hudson (1992), um levantamento de solos consiste em uma estratégia científica baseada nos conceitos de fatores de formação dos solos acoplados com relações solo-paisagem. Por outro lado Bui (2004) considera que um levantamento de solos consiste em um “sistema de conhecimentos” e que os mapas derivados desses conhecimentos são de fato representações estruturadas sobre a distribuição dos solos na paisagem.

A elaboração de um mapa em um levantamento convencional de solos envolve três etapas. Na primeira é realizada uma análise exploratória do local com os dados disponíveis, como por exemplo, fotografias aéreas, geologia, vegetação, entre outros. Num segundo momento faz-se fotointerpretação analisar a paisagem. Em seguida realiza-se uma checagem a campo para conferência com as fotos e, por último, a partir da interpretação dos dados de campo, é criado um modelo conceitual de solo-paisagem para inferir as variações espaciais do solo e, por último, o modelo conceitual é aplicado para deduzir a distribuição espacial dos solos no restante da área e assim delimitar as unidades de mapeamento de solos e a elaboração dos mapas (Hudson, 1992; Hash, 2008).

Igualmente aos levantamentos pedológicos convencionais, os estudos em mapeamento digital de solos (MDS) também baseia-se nas relações solo-paisagem. Entretanto, enquanto que nos levantamentos convencionais as relações solo-paisagem são estabelecidas de forma qualitativa, no mapeamento digital elas são avaliadas quantitativamente a partir de variáveis geomorfométricas derivadas de Modelo Digital de Elevação (MDE). A partir disso, Lagacherie e McBratney (2007) definiram o mapeamento digital de solos

(MDS) como sendo o modelo que se baseia na geração de sistemas de informações e que permitem estabelecer relações matemáticas entre variáveis ambientais (atributos do terreno) e classes de solos e, assim, predizer a distribuição espacial das classes de solos com significativa semelhança aos mapas gerados por meios convencionais.

O MDS começou na década de 1970 e acelerou-se com o advento de novas tecnologias de coleta, processamento e análise de informações espaciais, como os Modelos Digitais de Elevação (MDE), os Sistemas de Informação Geográfica (SIG), o Sistema de Posicionamento Global (GPS) e o sensoriamento remoto (McBratney et al., 2003; Rossiter, 2004). A abordagem do mapeamento digital de solos apresenta duas linhas de estudo, uma com enfoque no mapeamento de atributos do solo (teor de argila, teor de matéria orgânica, fósforo, etc..) e a outra linha com enfoque no mapeamento de classes de solos. Como o enfoque do nosso estudo relacionou-se à linha de mapeamento de classes buscou-se a realização de um levantamento de artigos publicados, tanto no Brasil como no exterior (em inglês) que abordassem essa linha de estudo, com a finalidade de evidenciar quais os principais modelos e as principais variáveis que estão sendo utilizados na realização desses estudos. Essa pesquisa baseou-se somente em artigos científicos. Não foram incluídos artigos que constam em capítulos de livros. A relação dos artigos mencionados encontram-se na Tabela 1

2.2. Modelos preditivos empregados no MDS

Entre as relações matemáticas mais utilizadas no Mapeamento Digital de classes de solos predomina a técnica de mineração de dados por árvore de decisão (AD), seguido pela técnica de regressão logística, lógica fuzzy e redes neurais artificiais.

Os primeiros estudos que demonstraram a eficiência do uso das ADs para o mapeamento preditivo de solo-paisagem em escala regional e sub-regional foi realizada por Moran e Bui (2001) e Scull et al. (2005), entretanto, a partir de 2011 constatou-se um crescente aumento na utilização dessa metodologia. Isso ocorre porque as ADs são ferramentas eficientes para a predição de ocorrência de solos pois permitem a extração do conhecimento pedológico a partir de informações pré-existentes e correlacioná-las

(Henderson et al., 2005). Isso é possível porque elas simulam o processo de abstração humana através de uma categorização hierárquica, obtendo regras similares a uma chave de classificação. Esse processo ocorre pois elas dividem recursivamente os dados para formar subconjuntos homogêneos, resultando em uma árvore hierárquica de regras de decisão (Behrens e Scholten, 2007). Na prática, ela aplica uma série de regras necessárias para produzir as classes finais desejadas ou para obter subconjuntos tão homogêneos quanto possíveis. A classe final constitui a folha e, da mesma forma que ocorre com uma árvore real, algumas classes (folhas) podem ser geradas com um número de divisões (nós) diferente de outras (Miller e Franklin, 2001). As vantagens de utilizar esse método é que geralmente são fáceis de entender, são capazes de processar um grande volume de dados e relacionar variáveis discretas e contínuas e validar os resultados através de métodos estatísticos, não são sensíveis à falta de dados e a inclusão de variáveis irrelevantes, possibilitam o uso de dados de diferentes tipos e fontes, como imagens de sensoriamento remoto, MDE, variáveis derivadas do MDE e mapas existentes (Miller e Franklin, 2001).

A utilização de regressões logísticas múltiplas (RLM) no MDS é possível pois geram um valor de probabilidade de ocorrência para cada classe de solo a ser mapeada (Giasson et al., 2006). Os valores de probabilidade de uma classe são comparados e aquele que tiver o maior valor é atribuído ao local. Essas respostas assumem a existência de uma relação linear entre a variável dependente e duas ou mais variáveis independentes. Essa metodologia foi utilizada para a determinação de classes de solos, primeiramente por Bailey et al. (2003) e mais recentemente por Giasson et al. (2006), Figueiredo et al. (2008) e Kempen et al. (2009).

A lógica fuzzy é caracterizada por um grau de pertinência que varia de 0 a 1, indicando um incremento contínuo (Zhu et al., 1996). Dessa forma a lógica difusa é capaz de incorporar a incerteza presente em diferentes tipos de dados. Essa técnica tem sido bastante explorada em estudos de mapeamento digital de solos pois possui a habilidade de capturar e representar a variação espacial de propriedades ou classes de solo (Zhu et al., 1996; Zhu et al., 2001; McMillan et al., 2007).

As redes neurais artificiais são algoritmos que tentam simular as informações de forma semelhante ao cérebro humano (Gershenfeld, 1999). Elas resolvem um problema desenvolvendo uma memória ou “aprendem” por meio de exemplos e depois aplicam essa memória para produzir uma saída quando um determinado conjunto de dados é fornecido (Qi & Zhu, 2003). O aprendizado é necessário para a obtenção de pesos que possibilitam a extração de informações de entrada para fazer a predição. Esse processo não gera modelos de fácil entendimento apesar de gerar mapas com boas acurácias (Kheir et al., 2010). Essa metodologia foi utilizada por Zhu (2000), Behrens et al. (2005) e Chagas et al. (2007), entre outros.

2.3. Variáveis utilizadas para o treinamento dos modelos

O modelo s.c.o.r.p.a.n utilizado como base no MDS foi proposto por McBratney et al. (2003) e podem ser treinados com uma grande quantidade de variáveis, dentre elas, variáveis do próprio solo, variáveis de clima, variáveis referentes aos organismos (incluindo cobertura vegetal, uso do solo, etc.), variáveis derivadas do relevo (elevação, declividade, etc.), material de origem, tempo e a posição espacial. Entre as variáveis utilizadas para o treinamento dos modelos nos estudos relacionados na Tabela 1 constata-se que as variáveis mais utilizadas estão relacionadas aos fatores de formação organismos, material de origem e, principalmente, ao relevo. Neste sentido ten Caten et al. (2012) enfatiza que a predominância da utilização de 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 relacionadas ao relevo que aparecem com maior frequência nos trabalhos são a declividade e a elevação. A declividade aparece em praticamente todos os trabalhos, enquanto a elevação, somente em raras exceções (Lagacherie et al., 2001; Moran e Bui, 2002; Nield et al., 2006; Hengl et al., 2007; Crivelenti et al., 2009; Yang et al., 2011; Cavazzi et al., 2013 e Adhikari et al., 2014) não é utilizada. As variáveis curvatura (do perfil ou planar), aspecto ou orientação e fluxos (comprimento e acúmulo) também aparecem com bastante frequência.

Os índices ou variáveis secundárias também aparecem como opção de treinamento dos modelos. Os mais comumente utilizados são o índice de umidade topográfico ou índice topográfico composto (Scull et al., 2005; Giasson et al., 2006; Nelson e Odeh, 2009, entre outros) e os índices de vegetação (Grinand et al., 2008; Nelson e Odeh, 2009; Odgers et al., 2011 e ten Caten et al., 2013).

Além das variáveis geomorfológicas e dos índices, alguns autores também utilizaram informações de geologia (Qi e Zhu, 2003; Behrens et al., 2005; Nolasco-Carvalho et al., 2009, entre outros), cobertura vegetal ou uso da terra (Zhu et al., 2001; Odgers et al., 2011; Kempen et al., 2012, entre outros), material de origem (Zhu et al., 1996; Lagacherie et al., 2001; Bui e Moran, 2001). Entretanto, apesar da variedade de variáveis utilizadas, as mais comuns são aquelas derivadas dos MDE, tornando evidente a importância dessa ferramenta. Isso é possível pois os MDE estão disponíveis para praticamente todas as regiões do mundo, com diferentes resoluções e, principalmente, de forma gratuita (Behrens et al., 2010). A seguir será realizada uma breve apresentação e discussão das principais características das variáveis derivadas do MDE e sua importância na formação dos solos.

2.3.1. Classificação e caracterização das variáveis geomorfológicas

As variáveis ambientais derivadas do relevo utilizadas para treinar os modelos no MDS são chamadas de variáveis geomorfológicas e são classificadas em duas categorias quanto ao método de cálculo e quanto a quantidade de pixels necessários para obtê-las. Quanto ao método de cálculo as variáveis são classificadas em primárias e secundárias (Moore et al., 1993). As variáveis primárias são aquelas derivadas diretamente do MDE, sendo elas: elevação do terreno, declividade, orientação, curvatura de perfil e planar, comprimento de fluxo, área de contribuição e acúmulo de fluxo. As secundárias são as variáveis que envolvem a combinação de variáveis primárias e, geralmente, são índices geomorfológicos, tais como o índice de umidade topográfica (IUT), índice de poder de escoamento, classificação das formas de relevo, entre outros. Quanto à quantidade de pixels utilizada para os cálculos podem ser chamadas de variáveis locais ou regionais (Olaya, 2009). As variáveis locais são aquelas que podem ser calculadas com uma pequena

quantidade de pixels (janela 3x3), como por exemplo, declividade e aspecto ou orientação. As regionais necessitam de áreas maiores para o cálculo pois definem os movimentos dos fluxos e necessitam das relações entre os pixels, como por exemplo o acúmulo de fluxo e a área de contribuição.

2.3.2 Variáveis primárias

Uma das variáveis primárias mais importante é a elevação do terreno. A elevação é importante pois está relacionada aos gradientes climáticos locais (Wilson e Gallant, 2000). Quando se trata da sua importância na formação dos solos a elevação é importante pois tem influência direta com a temperatura, a qual é muito importante para o acúmulo de matéria orgânica e, além disso, há uma tendência de ocorrerem maiores precipitações em maiores altitudes (Kämpf e Curi, 2012).

A declividade, outra variável bastante utilizada em MDS, tem grande importância na hidrologia e geomorfologia. A declividade afeta a velocidade dos fluxos superficial e sub-superficial da água e, assim, influencia o teor de água no solo, os processos erosivos e a formação dos solos, entre outros (Wilson e Gallant, 2000). Em função dos fluxos serem mais intensos em regiões que apresentam maiores declividades ou mais íngremes, é comum encontrar afloramentos de rochas e solos menos desenvolvidos como Neossolos Regolíticos e Litólicos (Kämpf e Curi, 2012).

A curvatura está relacionada à convexidade ou a concavidade da superfície do terreno em uma determinada direção (Behrens et al., 2010). Segundo Wilson e Gallant, (2000) as medidas de curvatura mais usualmente utilizadas são a curvatura de perfil e a curvatura planar. A curvatura de perfil mede a taxa de variação do gradiente potencial, sendo assim importante para a caracterização das mudanças da velocidade dos fluxos e o processo de transporte de sedimentos, enquanto a curvatura planar é a medida de convergência e divergência do relevo e, portanto, mede a propensão da água convergir para uma determinada direção.

A descrição da curvatura do terreno é importante pois está relacionada com a velocidade do escoamento, direção dos fluxos laterais, taxas de erosão e deposição, entre outros (Romano e Chirico, 2004). Como o movimento e a distribuição da água são as principais causas do deslocamento de partículas,

consequentemente, também são os principais fatores das diferenças entre solos (Kämpf e Curi, 2012). Como exemplo disso, em áreas com formatos convexos, nos quais o processo de erosão é mais intenso, há a ocorrência de solos menos desenvolvidos, enquanto que em regiões côncavas os fluxos saturados tendem a predominar e, assim sendo, ocorre o predomínio de solos com características de hidromorfismo.

As variáveis relacionadas aos fluxos descrevem o modo de movimentação da água. A direção de fluxo indica a direção do movimento da água, ou seja, a direção que o fluxo segue para sair de uma determinada célula/pixel (Wilson e Gallant, 2000). A partir da informação da direção do fluxo é possível obter o acúmulo de fluxo de cada célula, pois o mesmo é calculado em função da quantidade de células a montante que contribuem com o fluxo para uma determinada célula a jusante, assim sendo, representa a área de contribuição para uma determinada célula (Dobos et al., 2006), como exemplificado na Figura 1. Em relação ao acúmulo de fluxo, Gessler et al. (2000) constataram que juntamente com a curvatura e a declividade, foi a variável que teve uma significativa contribuição para a determinação da profundidade do horizonte A e da quantidade de matéria orgânica presente no mesmo.

O comprimento de fluxo calcula a distância a montante ou a jusante do trajeto de escoamento para cada célula. O comprimento de fluxo é muito utilizado para calcular o trajeto de escoamento mais longo dentro de uma bacia, ou seja, o tempo de concentração da mesma (Wilson e Gallant, 2000).

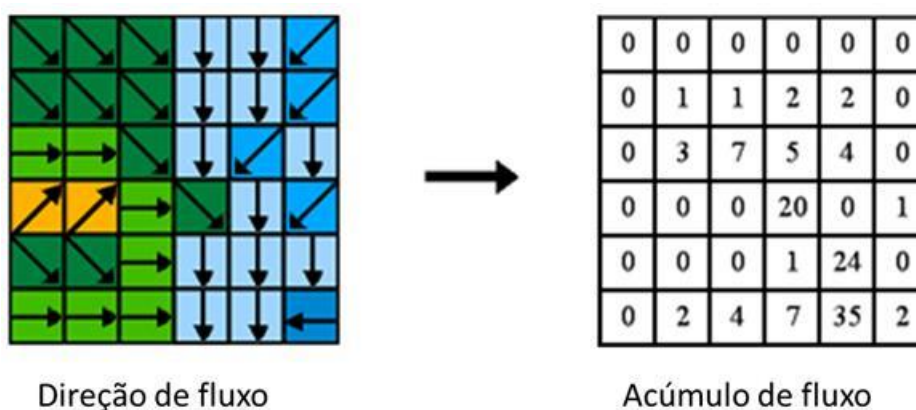


FIGURA 1. Esquema da direção e acúmulo de fluxo.

O aspecto ou orientação é a direção para a qual a normal à superfície do terreno está voltada (Wilson e Gallant, 2000). A orientação da superfície expõe ou oculta a superfície da terra afetando a interceptação de radiação solar, sendo assim influenciando na evaporação da água e no regime hídrico do solo.

2.3.3. Variáveis secundárias

Um dos índices mais conhecidos e mais usados em MDS é o índice de umidade topográfica (IUT) ou “topographic wetness index” (TWI). Esse índice foi desenvolvido por Kirkby (1975) e é utilizado para quantificar o efeito da topografia em processos hidrológicos, modelar a distribuição espacial das zonas de saturação superficial e determinar o conteúdo de água nas paisagens. Apesar de ser um índice amplamente utilizado, Hengl e Rossiter (2003) destacam que deve-se ter um cuidado em locais onde a declividade é zero, pois o cálculo falha, já que não é possível a divisão por um valor nulo.

O índice de posição topográfica (“topographic position index” – TPI) reflete a diferença de elevação de uma célula em relação a média de elevação das células vizinhas (Jenness, 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. Valores nulos podem indicar a presença de áreas planas ou áreas da meia encosta, sendo que, o que define ser um ou outro é a inclinação da célula. Valores negativos podem indicar que a célula está no fundo de um vale ou muito próximo de um vale. Em função dos valores de TPI, Jenness (2006) criou classes para classificar a declividade e chamou esse índice de classificação da posição da declividade (“slope position classification” – SPC). As classes de declividade criadas por Jenness foram: vale, declividade baixa, declividade plana, declividade média e cume, como pode ser observado na Figura 2. Essas classes foram separadas em duas escalas, uma utilizando uma pequena escala para o cálculo e uma utilizando uma escala maior. A utilização das classificações gerou um novo índice, a classificação das formas de relevo (“landform classification” – LC). Jenness (2006) classificou as formas de relevo como colina, topo de colina, regiões planas, pequenas colinas em regiões planas.

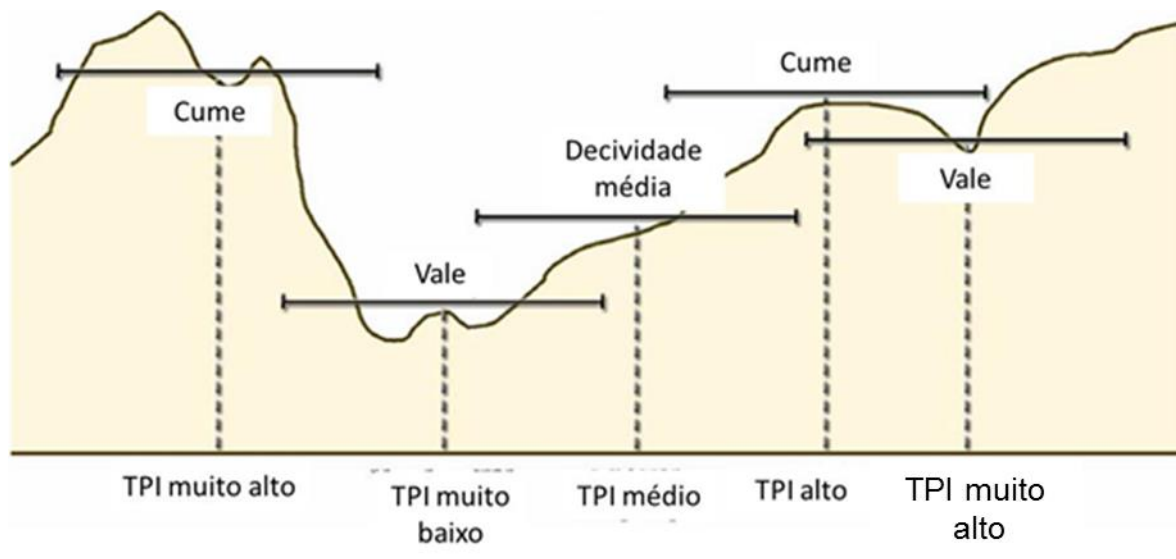


FIGURA 2. Esquema das classes de declividade.
adaptado de Jenness (2006)

2.4. Avaliação da acurácia dos mapas gerados por MDS.

Após a aplicação dos modelos e geração dos mapas faz-se necessário a avaliação da qualidade dos mapas gerados. Nos estudos relacionados na Tabela 1 constatou-se a utilização de dois métodos de avaliação da acurácia. Os métodos utilizados foram a matriz de erro (Congalton, 2001) e o índice *Kappa* (Cohen, 1960).

A matriz de erro foi a opção mais utilizada. Na matriz de erro os resultados do mapa predito são relacionados com os resultados do mapa original, na qual, as colunas representam os dados do mapa predito e as linhas os dados do mapa original (Congalton, 2001). A partir da matriz é possível extrair os resultados de erros de inclusão, erros de omissão e a exatidão geral. Os resultados de exatidão geral são obtidos a partir da diagonal principal da tabela, na qual a classe estimada e a de referência são concordantes (Foody, 2002). O índice *Kappa*, também leva em consideração o grau de concordância entre os mapas, entretanto, elimina os resultados que ocorreram puramente ao acaso (Elnaggar, 2007). Esse índice é bastante utilizado, principalmente em estudos relacionados ao sensoriamento remoto.

Entre os estudos analisados (dados não apresentados) observou-se uma ampla variação de resultados obtidos. O menor valor de acurácia geral foi observado por ten Caten et al. (2011a). O valor encontrado por esses autores foi de 39,3% quando da realização da extrapolação do modelo a bacia inteira

utilizando uma pequena área de referência e o maior valor foi de 87,3% por Dobos et al. (2000). Entretanto na grande maioria dos estudos os valores de acurácia geral obtidos encontram-se na faixa entre 60% e 70%.

Com relação aos trabalhos que apresentaram o índice Kappa como forma de avaliação da qualidade dos mapas os menores valores foram observados por Giasson et al. (2006) de 0,36 e os maiores de 0,73 por Chagas et al. (2010), entretanto, os valores mais frequentemente encontrados se encontram na faixa de 0,4 a 0,5.

Sendo assim, com base nos artigos relacionados na Tabela 1, pode-se perceber a tendência do uso de AD, com as variáveis altimetria, declividade, aspecto, fluxos (comprimento, direção e acúmulo) e distância de rios e a avaliação dos mapas gerados, principalmente, por matriz de erro. Com base nisso, essa base metodológica para estruturar os estudos dessa tese, na qual são avaliados alguns aspectos metodológicos que ainda necessitam de maiores esclarecimento.

TABELA 1. Relação de artigos referentes ao mapeamento digital de classes de solos

Autores	Ano	Local	Metodologia	Variáveis
Zhu et al.	1996	EUA	lógica fuzzy	material de origem, elevação, aspecto, cobertura vegetal e curvatura
Thomas et al.	1999	França	Função discriminante	elevação, declividade, aspecto, curvatura de perfil, curvatura planar e distância de rios
Zhu	2000	EUA	redes neurais	elevação, declividade, aspecto (seno e cosseno) cobertura vegetal, curvatura de perfil e geologia
Dobos et al.	2000	Hungria	Extração de características com análise discriminante	Radiometro avançado de alta resolução (AVHRR), elevação, declividade, aspecto e curvatura
Park et al.	2001	EUA		gradiente da declividade, aspecto, curvatura planar, área de contribuição, elevação, curvatura de perfil, índice de umidade, índice topográfico combinado e curvatura da superfície
Lagacherie et al.	2001	França	Clapas	material de origem, gradiente de declividade e altura relativa dos rios
Zhu et al.	2001	EUA	Lógica fuzzy	elevação, gradiente de declividade, orientação da declividade, curvatura de perfil, curvatura planar e cobertura florestal
Bui e Moran	2001	Austrália	Árvore de decisão	litologia, elevação, relevo, energia de transporte e Landsat (4 bandas)
Moran e Bui	2002	Austrália	Árvore de decisão	log da área de contribuição, distância dos rios, distância das colinas, aspecto, declividade, curvatura e Landsat (bandas 1 e 4)
Bailey et al.	2003	Inglaterra	regressão logística	altura relativa, aspecto e declividade

Continua....

Qi e Zhu	2003	EUA	Redes neurais artificiais, árvore de decisão, Naïve Bayes,	elevação, gradiente de declividade, curvatura planar e geologia
Hengl e Rossiter	2003	Croácia	Componentes principais	Declividade, elevação, perfil da curvatura, curvatura planar, índice topográfico composto, profundidade do lençol freático, acúmulo de fluxo reflectância e distância dos rios
Bui e Moran	2004	Austrália	Árvore de decisão	relevo, elevação relativa, posição da declividade, área de contribuição, energia de transporte, distância dos canais e distância de rios
Zhou et al.	2004	China	Árvore de decisão e modelo preditivo de Bayesian	material de origem, uso da terra, elevação, declividade, aspecto, perfil da curvatura, curvatura planar e área de contribuição
Qi	2004	EUA	Árvore de decisão	elevação, gradiente da declividade, curvatura planar curvatura de perfil e geologia
Scull et al.	2005	EUA	Árvore de classificação/decisão	Landsat TM, temperatura máxima de julho, temperatura mínima de janeiro, precipitação de verão e de inverno, Landsat MSS, índice de umidade topográfica, radiação solar potencial, declividade, elevação, formas de relevo(várias) e vegetação e geologia
Behrens et al.	2005	Alemanha	Redes neurais artificiais	geologia, uso da terra e 69 atributos de terreno
Nield et al.	2006	Canadá	Diferenças de bandas Landsat	direção do fluxos, percurso de escoamento e Landsat 7 ETM
Giasson et al.	2006	Brasil	Regressões logísticas múltiplas	elevação, gradiente de declividade, curvatura de perfil, curvatura planar, curvatura(combinação de curvatura de perfil e planar), direção de fluxo, acúmulo de fluxo, comprimento de fluxo, capacidade de transporte de sedimentos e índice de umidade topográfica

Continua....

Smith et al.	2006	EUA	Abordagem SoLIM	gradiente de declividade, elevação, curvatura de contorno, curvatura de perfil, fuzzy narrow ridge e fuzzy broad ridge
Minasny e McBratney	2007	Austrália	Árvores de decisão	elevação, declividade, aspecto, índice de umidade topográfica, índice normalizado da diferença de vegetação, taxa de aprimoramento do solo e radios gamma de potássio
Hengl et. al.	2007	Iran	geoestatística	declividade, curvatura média, índice de umidade topográfica, radiação insidente direta, percentis de diferença de índices e índice de vegetação
MacMillan et al.	2007	Canadá	Lógica fuzzy	log da área de contribuição, índice de umidade de Quinn, declividade, aspecto, percentagem de mudança na elevação, mudança absoluta vertical na elevação, distância horizontal, distância vertical, distância horizontal e bandas 3 e 5 Landsat.
Chagas et al.,	2007	Brasil	Redes Neurais artificiais	elevação, declividade, aspecto, curvatura, índice topográfico combinado, geologia, índice de vegetação, índice de minerais de argila, índice de óxido de ferro
Figueiredo et al.	2008	Brasil	Regressões logísticas múltiplas	Elevação, declividade, curvatura, curvatura planar, perfil da curvatura, direção do escoamento, acumulação do escoamento, comprimento do escoamento, índice de umidade topográfica e distância dos rios
Grinand et al.	2008	França	Árvores de decisão	elevação, declividade, curvatura de perfil, curvatura planar, índice topográfico composto, distancia hidrológica relativa ao rio mais próximo, altura relativa ao rio mais próximo, todas as bandas ETM, Índice de vegetação por diferença normalizada (NDVI), geologia e cobertura vegetal

Continua....

Schmidt et al.	2008	Alemanha	Seleção instantânea e árvores de classificação/decisão	geologia, uso da terra e 69 atributos de terreno (ver Behrens et al., 2005)
Sorokina e Kozlov	2009	Rússia	Lógica fuzzy	elevação relativa, gradiente de declividade, curvatura da declividade, área de contribuição, índices de conteúdo de umidade, atividade de erosão, insolação solar anual, índices de vegetação (albedo, biomassa, produtividade da vegetação, reservas de água e densidade da vegetação do dossel).
Kempen et al.	2009	Holanda	Regressão logística multinomial	elevação, elevação relativa, água subterrânea, histórico da cobertura vegetal, paleografia e geomorfologia.
Nolasco-Carvalho et al.	2009	Brasil	Inferência fuzzy	Elevação, declividade, geologia e vegetação
Nelson e Odeh	2009	Austrália	Algoritmo genético e árvore de classificação	elevação, curvatura, curvatura de perfil, curvatura planar, aspecto, declividade, índice de planicidade, índice topográfico de umidade, classificação das formas de relevo de Hammond , magnetic anomaly, rádios gamma, uso da terra e NDVI
Crivelenti et al.	2009	Brasil	Árvores de decisão	curvatura em planta, curvatura em perfil, área de contribuição da bacia, declividade, distancia diagonal da drenagem e geologia
Chagas et al.	2010	Brasil	Redes neurais artificiais	elevação, declividade, aspecto, plano de curvatura e índice topográfico combinado

Continua....

Behrens et al.	2010	Alemanha	Componentes principais e ANOVA	elevação, declividade média, declividade mais íngreme, aspecto, curvatura média, curvatura de perfil, curvatura horizontal, curvatura mínima, curvatura máxima, relative curvatura de perfil, curvatura horizontal relativa, rugosidade, área de contribuição da bacia, distância projetada do rio, elevação local, posição relativa da encosta, índice topográfico composto e fator LS-USLE
Coelho et al.	2010	Brasil	Regressões Logísticas Múltiplas Multinomiais, Bayes e árvores de decisão	Elevação, declividade, curvatura planar, perfil de curvatura, curvatura, direção do fluxo, acúmulo de fluxo e índice de umidade topográfica
ten Caten et al.	2011a	Brasil	Regressão Logística Múltipla	elevação, declividade, distancia da drenagem, curvatuna planar, curvatura do perfil, radiação relativa disponível, logaritmo natural da área de contribuição, índice de umidade topográfica e capacidade de transporte de sedimentos
ten Caten et al.	2011b	Brasil	Componentes principais e Regressão Logística Múltipla	elevação, declividade, distancia da drenagem, curvatuna planar, curvatura do perfil, radiação relativa disponível, logaritmo natural da área de contribuição, índice de umidade topográfica e capacidade de transporte de sedimentos
ten Caten et al.	2011c	Brasil	Regressão Logística Múltipla	elevação, declividade, distancia da drenagem, curvatuna planar, curvatura do perfil, radiação relativa disponível, logaritmo natural da área de contribuição, índice de umidade topográfica e capacidade de transporte de sedimentos
Carvalho Junior et al.	2011	Brasil	Redes Neurais artificiais	elevação, declividade, curvatura planar, curvatura de perfil, curvatura, aspecto, acumulo de fluxo, direção de fluxo, distancia dos rios e elevação relativa

Continua....

Yang et al.	2011	Canadá	agrupamento fuzzy c-means	material de origem, gradiente da declividade, curvatura planar e índice de umidade topográfica
Odgers et al.	2011	Austrália	Regressão-kriging	Uso da terra, elevação, declividade, aspecto, cosseno do aspecto, seno do aspecto, índice topográfico composto, curvatura planar, curvatura de perfil, Landsat (bandas 8), NDVI, relevo, distância de rios, K, U, Th e a relação entre eles
Giasson et al.	2011	Brasil	Árvore de decisão	elevação, declividade, curvatura planar, perfil de curvatura, curvatura, direção do fluxo, acúmulo de fluxo, índice de umidade topográfica e índice de poder de córrego
ten Caten et al.	2012	Brasil	Árvores de decisão	elevação, declividade, índice de umidade topográfica, capacidade de transporte de sedimentos, curvatura planar, perfil da curvatura, índice de rugosidade do terreno,
Sarmiento et al.	2012	Brasil	Árvores de decisão e redes neurais	elevação, declividade, orientação, perfil de curvatura, acúmulo de fluxo, direção de fluxo e distância de rios
Kempen et al.	2012	Holanda	Modelo multinomial logit	elevação relativa, mudança de elevação, cobertura vegetal em 1900, cobertura vegetal atual, mapa de paleogeografia
ten Caten et al.	2013	Brasil	Árvores de decisão	elevação, declividade, distancia da drenagem, curvatura planar, curvatura do perfil, índice de umidade topográfica, rugosidade e NDVI
Cavazzi et al.	2013	Irlanda	Árvores de decisão	declividade, aspecto, curvatura, curvatura planar, curvatura de perfil, altura da declividade, profundidade dos vales, altura normalizada, altura standarizada, declividade na posição média e índice de convergência

Continua...

Giasson et al.	2013	Brasil	Árvores de decisão	elevação, declividade, curvatura, comprimento do fluxo a jusante, comprimento do fluxo a montante, acúmulo de fluxo, índice de umidade topográfica e distância de rios
Adhikari et al.	2014	Dinamarca	Árvores de decisão	orientação da declividade, quantidade de argila, luz solar incidente, distância de rios, elevação, acúmulo de fluxo, geologia, regiões geológicas, formas de relevo, uso da terra, declividade média, mid-slope position, multi-resolução do índice de planicidade, índice de umidade SAGA , profundidade dos vales e áreas úmidas
Teske et al.	2014	Brasil	Árvores de decisão	elevação, declividade, curvatura, curvatura em perfil, curvatura planar, direção de fluxo, acúmulo de fluxo e orientação das vertentes, índice de umidade topográfica, índice de convergência topográfica e índice de posição topográfica
Höfig et al.	2014	Brasil	Árvores de decisão	elevação, declividade, curvatura, orientação das vertentes, distância dos rios, direção de fluxo, acúmulo de fluxo e índice de umidade topográfica
Teske et al.	2015	Brasil	Árvores de decisão	elevação, declividade, direção do fluxo, acúmulo do fluxo, comprimento do fluxo, curvatura, índice de umidade topográfica e distância de rios

3. MATERIAL E MÉTODOS

3.1. Localização das áreas de estudo

Para a realização deste trabalho foram utilizadas três bacias hidrográficas situadas em áreas diferenciadas ecologicamente, sendo duas representativas do Planalto (bacia do rio Lageado Grande e do Santo Cristo) e uma da Depressão Central (bacia do Arroio Portão) do Estado do Rio Grande do Sul. A localização dessas bacias são apresentadas na Figura 3.

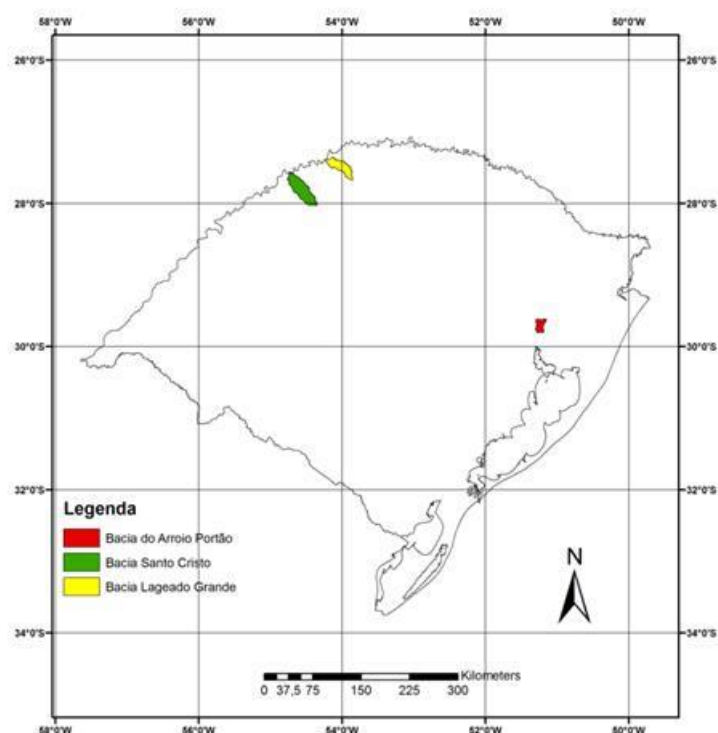


FIGURA 3. Áreas das bacias hidrográficas utilizadas no estudo.

3.1.1. Bacias do Lageado Grande e Santo Cristo

A bacia do rio Lageado Grande situa-se na porção noroeste do Estado do Rio Grande do Sul, inserida na Bacia Hidrográfica U30; ocupa uma área

equivalente a 531 km², abrangendo áreas dos municípios de Campo Novo (11,27%), Sede Nova (26,51%), Humaitá (52,24%), Bom Progresso (20,89%), Três Passos (37,96%), Tiradentes do Sul (37,10%) e Crissiumal (49,78%). Destes municípios, tem sua sede localizada dentro dos limites da microbacia, os seguintes: Sede Nova, Humaitá, Bom Progresso, Três Passos, Tiradentes do Sul e Crissiumal.

A bacia do rio Santo Cristo situa-se na porção noroeste do Estado do Rio Grande do Sul, inserida na Bacia Hidrográfica U30; ocupa uma área equivalente a 837 km², abrangendo áreas dos municípios de Giruá (12,90%), Santa Rosa (64,11%), Santo Cristo (44,11%), Tuparendi (54,92%), Porto Mauá (34,62%), Senador Salgado Filho (15,90%) e Alecrim (27,47%). Destes municípios, Giruá, Tuparendi e Alecrim tem sua sede localizada sobre o divisor de águas da microbacia.

O ambiente geológico dominante nas bacias é constituído por rochas vulcânicas da Formação Serra Geral, principalmente na forma de rochas basálticas. Esta litologia comum, através do intemperismo, forneceu o material de origem para a formação dos solos na região. O relevo da região está relacionado a litologia e a tectônica regional, distinguindo-se três fases principais: relevo plano a suavemente ondulado, relevo ondulado e relevo fortemente ondulado a montanhoso. A variação da declividade do terreno está relacionada a inserção dos cursos de água componentes da rede hidrográfica da microbacia. Considerando que as rochas basálticas são o material de origem comum a todos os solos, o principal fator diferenciador entre os tipos de solo é o relevo, condicionado aos fluxos de água e a drenagem, os processos erosivos e a acumulação de materiais.

Os mapas de solos utilizados encontram-se na escala de 1:50.000 e faz parte dos levantamentos pedológicos e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da bacia do Rio Lageado Grande (Kämpf et al., 2004a) e dos levantamentos pedológicos e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da bacia do Rio Santo Cristo (Kämpf et al., 2004b).

A bacia é composta por 15 unidades de mapeamento de solos (UMS), sendo a composição das mesmas apresentada na Tabela 2 e a distribuição destas e da altitude são apresentadas nas Figuras 4a e 4b, respectivamente. A

bacia do Santo Cristo é composta por 10 UMS e a relação da composição de cada UMS é apresentada na Tabela 3 e a distribuição das mesmas na área da bacia e da sua respectiva altitude são apresentados nas Figuras 5a e 5b, respectivamente.

TABELA 2. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Lageado Grande: simbologia usada no mapa de solos e a área e representatividade de ocorrência da UM

UM	Composição da Unidade de Mapeamento	Lageado Grande	
		Área km ²	Área %
LV1	Latossolo Vermelho	144,9	27,2
LV2	Associação Latossolo Vermelho e Neossolo Regolítico	31,2	5,9
NV1	Nitossolo Vermelho	2,4	0,4
NV2	Associação Nitossolo Vermelho e Neossolo Regolítico	9	1,7
M1	Chernossolo Argilúvico	3,6	0,7
M2	Associação Chernossolo e Neossolo Regolítico	3,9	0,7
RR1	Neossolo Regolítico	161,8	30,4
RR2	Associação Neossolo Regolítico e Neossolo Litólico, relevo forte ondulado	125,9	23,7
RR3	Associação Neossolo Regolítico e Latossolo Vermelho	28,1	5,3
RR4	Associação Neossolo Regolítico e Chernossolo Argilúvico	4,7	0,9
RR5	Associação Neossolo Regolítico e Cambissolo e Nitossolo Vermelho	2,7	0,5
RR6	Associação Neossolo Regolítico + afloramento rochoso	2,6	0,5
RL	Associação Neossolo Litólico e Neossolo Regolítico	2,2	0,4
G	Gleissolos Háplico e Gleissolo Melânico	8,7	1,6
Total		531	100

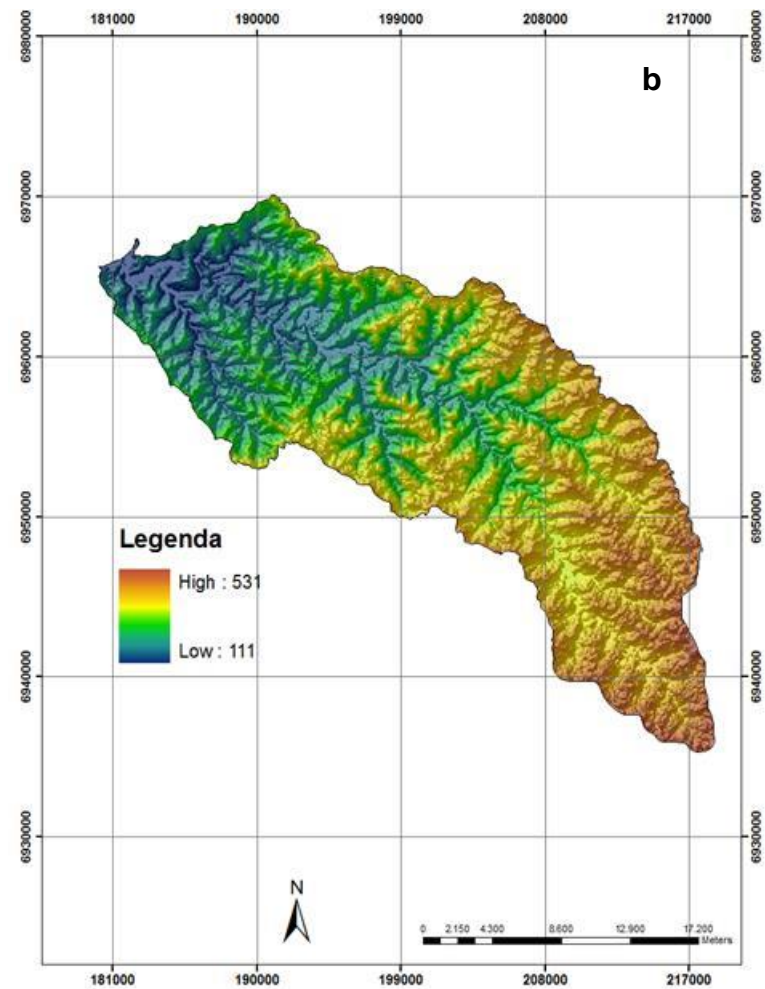
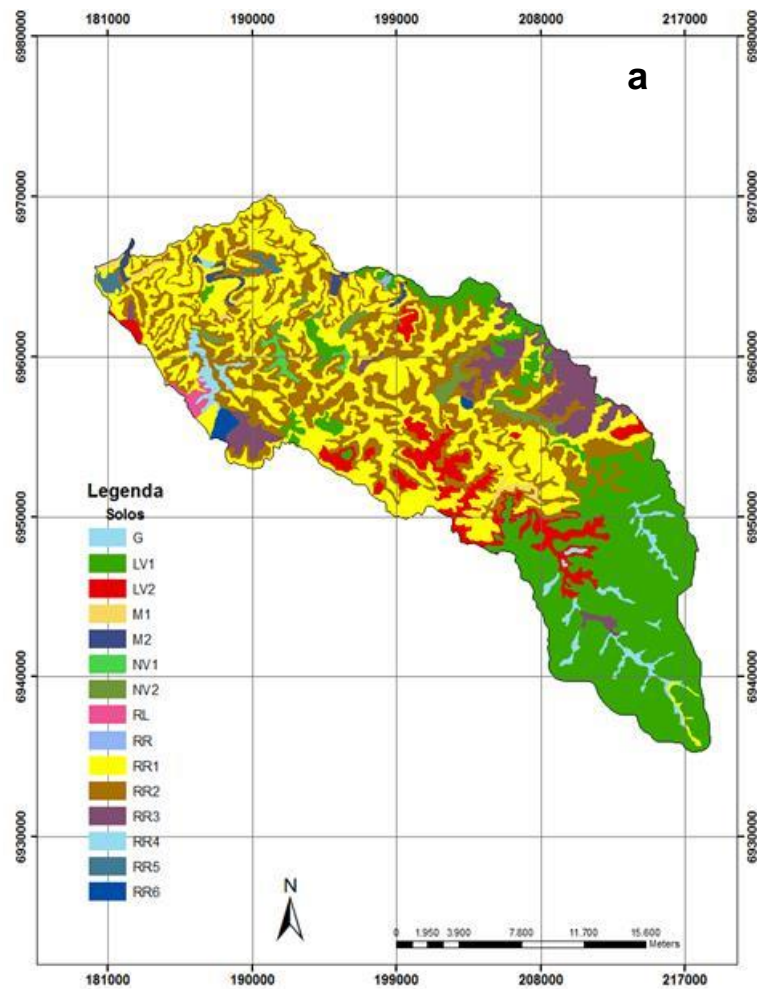


FIGURA 4. Mapa de solos (a) e mapa de altitude (b) da bacia do Lageado Grande com suas respectivas legendas.

TABELA 3. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Santo Cristo: simbologia usada no mapa de solos e a área e representatividade de ocorrência da UM

UM	Composição da Unidade de Mapeamento	Área	Área
	Santo Cristo	km²	%
LV1	Latossolo Vermelho	319,1	38,13
LV2	Associação Latossolo Vermelho e Neossolo Regolítico	66,3	7,92
M	Chernossolo Háplico	1,7	0,2
RL	Neossolo Litólico	0,01	0,01
RR1	Associação Neossolo Regolítico e Cambissolo Háplico	291,1	34,79
RR2	Complexo Neossolo Regolítico e Neossolo Litólico	69	8,24
RR3	Associação Neossolo Regolítico e Latossolo Vermelho	65,6	7,84
RR4	Associação Neossolo Regolítico e Neossolo Litólico	0,3	0,03
RR5	Associação Neossolos Regolíticos e Cambissolo Háplico e Latossolo Vermelho	3,2	0,38
G	Gleissolo Háplico e Gleissolo Melânico	20,6	2,46
Total		837	100

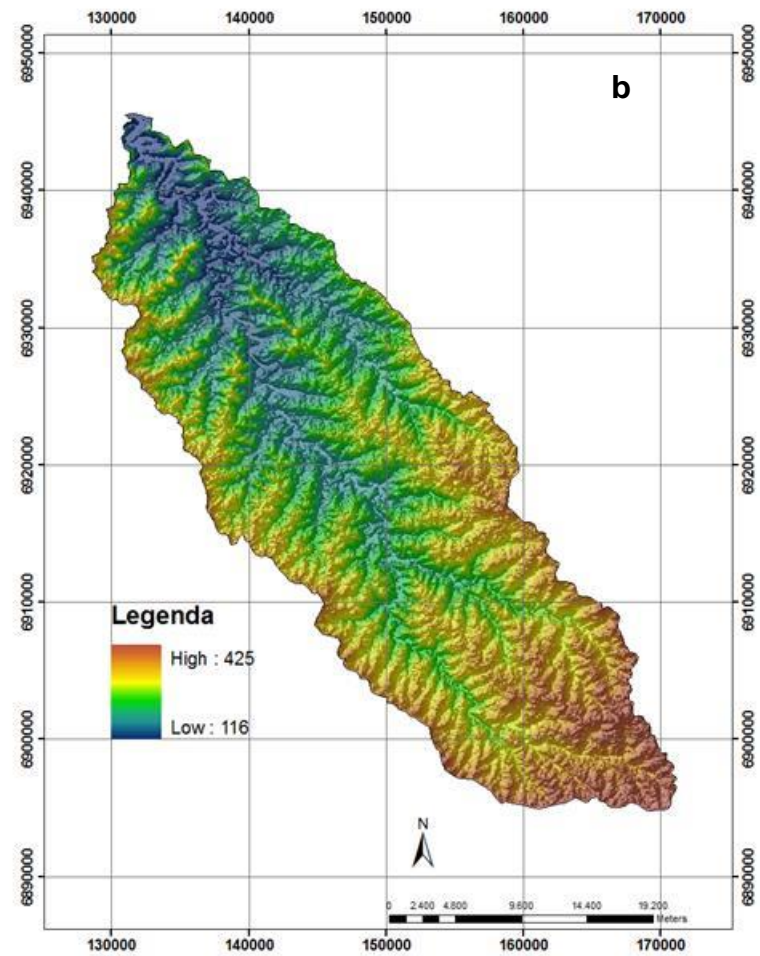
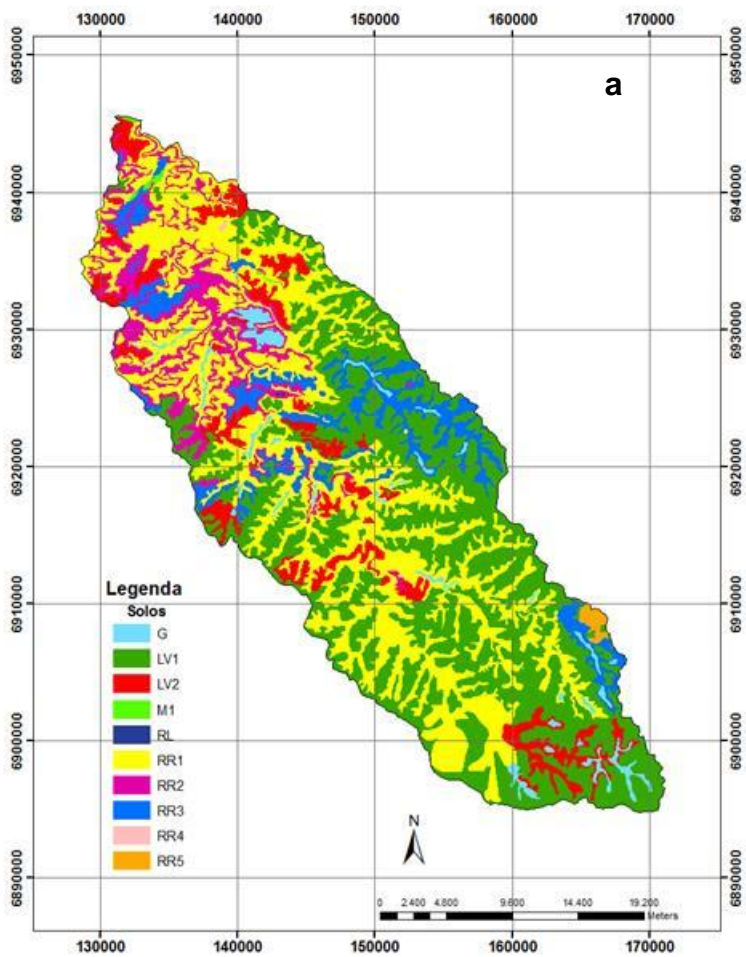


FIGURA 5. Mapa de solos (a) e mapa de relevo (b) da bacia do Santo Cristo com suas respectivas legendas.

3.1.3. Bacia do Arroio Portão

A área da bacia do Arroio Portão, compreende aproximadamente 221,6 Km² abrangendo parte dos municípios de Estância Velha, Portão, Ivoti, São Leopoldo e Canoas.

Na área ocorrem as formações geológicas Botucatu, Rosário do Sul, Serra Geral e os aluviões. A formação Botucatu é constituída por arenitos, apresenta estratificações eólicas e cores variando do róseo ao vermelho. A formação Rosário do Sul compreende faces fluvial com arenitos vermelhos com estratificação cruzada e intercalações de siltitos avermelhados. A formação Serra Geral compreende a zona de escarpas e platôs, sendo constituída por rochas vulcânicas, diques e sills de diabásio associado. Os aluviões são sedimentos atuais e sub-atuais depositados em planícies de inundação e barras, sendo constituídos basicamente por cascalhos, areia, silte e argila.

A área compreende a região fisiográfica da Depressão Central e a de transição para a Serra Geral. Nessa área existem três formas de relevo predominante: áreas baixas e planas, ao longo dos arroios e rios, formando extensas várzeas ao sul e ao sudeste da área; áreas constituídas por relevo suavemente ondulado a ondulado, compreendendo a maior parte da área onde predominam os arenitos (região mais ao centro da área) e zona escarpada, com relevo acidentado, compreendendo a zona de transição entre os arenitos e a formação Serra Geral, ao norte e nordeste da área. As informações referentes aos solos desta bacia hidrográfica são provenientes do relatório de levantamento de solos elaborado por Klamt et al. (1992). O levantamento de solos foi realizado na escala de 1:50.000, sendo o mesmo composto por 9 UMS. A composição das mesmas são apresentadas na Tabela 4 e a distribuição das mesmas e o mapa de altitude são apresentados nas Figuras 6a e 6b, respectivamente.

TABELA 4. Legenda de identificação das Unidades de Mapeamento (UM) da bacia Santo Cristo e da bacia do Arroio Portão: composição taxonômica das UM, área e representatividade de ocorrência da UM

UM	Composição da Unidade de Mapeamento	Área, km ²	Área, %
Arroio Portão			
CHa	Cambissolo Húmico	0,9	0,4
PVd	Argissolo Vermelho	47,6	21,2
PVa	Argissolo vermelho-amarelo	35,1	15,7
MXo/RLe	Chernossolo Háplico e Neossolo litólico	13,7	6,0
SXd	Planossolo Háplico	1,2	0,6
Aluvial	Solos hidromórficos	0,6	0,3
CXf/RLe	Cambissolo Háplico e Neossolo litólico	12,3	5,8
GMd	Gleissolo Melânico	17,68	8,4
PVa/GMe	Argissolo Vermelho e Gleissolo Melânico	92,6	41,3
Total		221,6	100

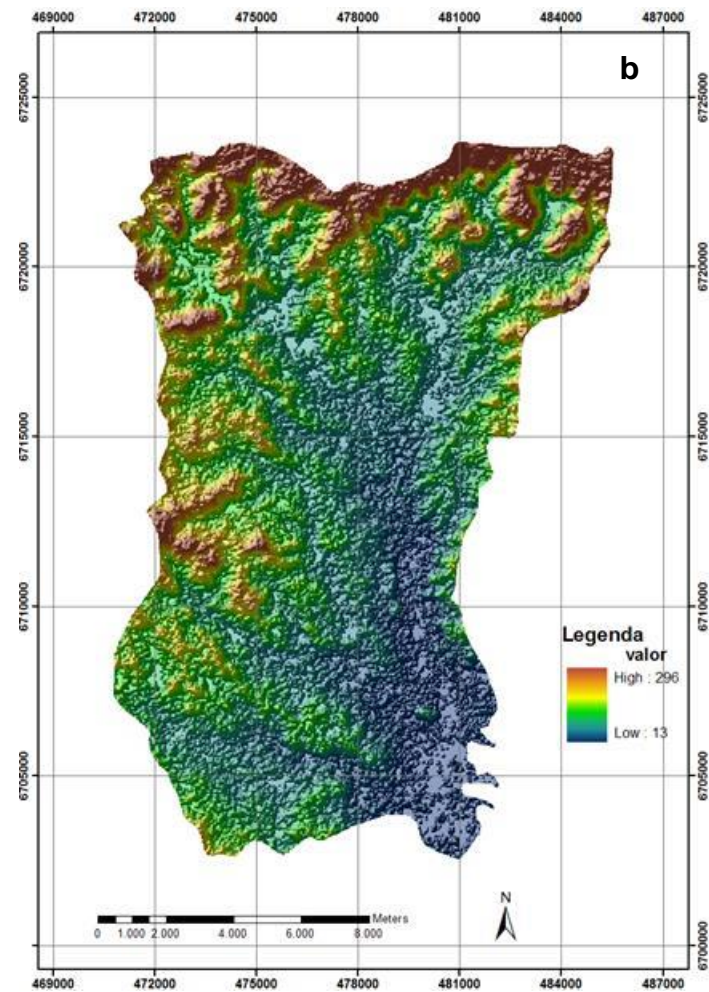
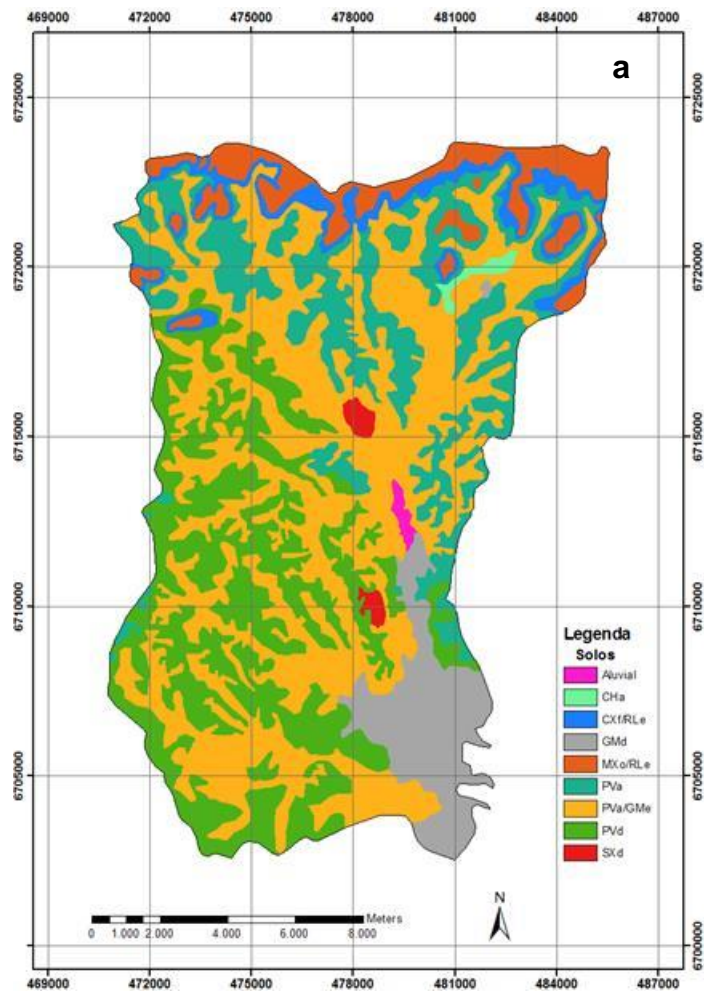


FIGURA 6. Mapa de solos (a) e mapa de relevo (b) da bacia do Arroio Portão com suas respectivas legendas.

3.2. Procedimentos gerais para a obtenção das variáveis geomorfométricas e aplicação dos modelos

Uma parte fundamental no MDS é a seleção das fontes de informação que irão alimentar o Sistema de Informação Geográfica (SIG), bem como as ferramentas computacionais que foram empregadas. Para a realização deste estudo foi utilizado o SIG ArcGis 9.3 (ESRI, 2006) e para a realização dos cálculos das variáveis geomorfométricas utilizou-se o modelo digital de elevação (MDE) ASTER-GDEM com resolução espacial de 30 metros (ABRAMS et al., 1999).

A partir do MDE, utilizando-se ferrametas do *Spatial analyst tools* foram gerados mapas de variáveis preditoras dos solos na paisagem. Para os estudos 1 e 3 foram utilizadas as variáveis altimetria (próprio MDE), declividade (função *slope*), acúmulo do fluxo (função *flow accumulation*), comprimento do fluxo (função *flow length*), curvatura do declive (função *curvature*) e índice de umidade topográfica (função *Topographic Convergence Index*) (BEVEN & KIRKBY, 1979) da ferramenta *Topography Tools*. Além dessas variáveis calculadas a partir do MDE também foi calculada a variável distância dos rios (função *euclidean distance*) que, para tanto foi utilizado o arquivo vetorial de hidrografia da base contínua do Rio Grande do Sul (HASENACK & WEBER, 2010). No estudo 2, além dessas foram utilizados também as variáveis aspecto (função *Aspect*) classificação da posição do declive (função *slope position classification*), índice de posição topográfica (função *topographic position index*) e classificação das formas de relevo (função *landform classification*). Todos os mapas citados foram gerados com resolução espacial de 30 metros.

Após o cálculo de todas as variáveis, em função de que cada uma das variáveis apresentam escalas de valores e unidades de medidas diferentes, afim de colocá-las todas numa mesma grandeza, todas as variáveis passaram pelo processo de normalização estatística utilizando a ferramenta *raster calculator*. Dessa maneira todas as variáveis ficaram com média zero e desvio padrão um. Para a realização de tal cálculo utilizou-se a seguinte fórmula.

$$Vn = \frac{(Vo - \mu)}{\sigma},$$

onde V_n é a variável normalizada, V_o a variável original (mapa), μ a média das variáveis e σ o desvio padrão.

Após a realização da normalização das variáveis procedeu-se a elaboração dos arquivos de amostragem. Para tal utilizou-se a ferramenta *create random points* na densidade específica para cada trabalho e realizada de forma simples e aleatória pelo próprio programa. A partir dessas amostragens as informações de todas as variáveis preditoras e das UMS ocorrente foram coletadas utilizando-se função *Sample*. Esses dados foram exportados do SIG e, posteriormente organizados em excel, de modo que, cada coluna correspondesse a uma variável e cada linha correspondesse a um ponto de amostragem. Esses dados foram exportados para proceder ao treinamento dos modelos com o algoritmo Simple Cart no programa Weka 3.6.3 (Hall et al., 2009).

A escolha pelo algoritmo Simple cart deveu-se ao fato que, dos algoritmos testados (J48 e Simple cart) para as bacias em análise, foi o que resultou nos melhores valores de acurácia geral, concomitantemente, com a geração de menores ADs. O teste utilizado para a construção do modelo foi *Cross Validation* (10 folds) e para a avaliação da qualidade dos mapas estimados usou-se como referência a comparação de correspondências entre o mapa gerado pelo modelo com o mapa original de solos. Essa metodologia empregada é baseada na construção de matrizes de erro (Congalton, 1991).

Por meio da matriz de erros é possível calcular a acurácia geral (AG), que representa a concordância entre o mapa de referência e a aplicação do modelo na geração de um mapa de solos usando MDS, a acurácia de usuário (AU) e a acurácia do mapeador (AM). A AM expressa o quanto do mapeado pelo MDS que também foi mapeado da mesma forma no mapa convencional de solos. A AU expressa a chance de uma UM mapeada na MDS também ter sido mapeada naquele pixel como sendo a mesma UM no mapa convencional de solos.

Os demais detalhes, específicos de cada estudo, serão apresentadas no material e métodos dos mesmos.

4 CAPÍTULO I. SELEÇÃO DE DENSIDADE DE AMOSTRAGEM PARA CALIBRAÇÃO DE MODELOS DE ÁRVORE DE DECISÃO PARA O MAPEAMENTO DIGITAL DE CLASSES DE SOLOS

4.1 Introdução

O mapeamento digital do solo (MDS) é definido como a criação e população de sistemas de informação pedológica espacial através do uso de métodos observacionais de campo e laboratório, acoplados a sistemas de inferência pedológica espaciais e não-espaciais (Lagacherie e McBratney, 2007). A sua utilização com base em sistemas de informação geográfica (SIG), estatística e pedologia está aumentando nas últimas décadas devido ao aumento de fontes de dados numéricos, tais como aqueles fornecidos pelos modelos digitais de elevação da superfície do solo (MDE), combinado com o desenvolvimento de novas ferramentas de processamento de informação (McBratney et al., 2003). A predição de tipos de solos na paisagem a partir desta técnica consiste na utilização de modelos matemáticos que conseguem descrever as relações entre os fatores de formação do solo e suas ocorrências na paisagem. Dentre esses modelos de predição, os que vêm ganhando destaque são os algoritmos de aprendizagem de máquinas. Neste sentido Qi & Zhu (2003), comparando três desses algoritmos, concluíram que árvore de decisão (AD) possuem uma maior capacidade de predição de atributos ou tipos de solos. Contudo, apesar de estudos de Moran e Bui (2002), Grinand et al., (2008), Giasson et al., (2011), entre outros, indicarem que os modelos de ADs são adequados para uso no MDS, algumas procedimentos metodológicos ainda precisam ser definidos. Dentre eles está a densidade de pontos amostrais nos quais são coletados informações sobre as variáveis ambientais nos mapas de referência e que são usados para o treinamento dos modelos preditores (McBratney et al., 2003).

A seleção da densidade de amostragem é um aspecto importante, pois o número de amostras usadas no treinamento dos modelos pode afetar significativamente a capacidade de predição dos algoritmos e sua acurácia (Hjort & Marmion, 2008). Neste sentido, Zhu (2000) sugere adotar como número mínimo de amostras de pelo menos 30 vezes o número de unidades de mapeamento de solos existentes ou a serem preditas. Grinand et al. (2008), usando mapa na escala 1:250.000, apontaram que a densidade amostral sobre o mapa de solos referente a um ponto amostral a cada 10 ha no campo foi suficiente para captar a variabilidade de tipos de solos. Da mesma forma, em estudos realizados em MDS no Brasil recentemente há grandes diferenças quanto à densidade de amostragem sobre mapas de referência que tem sido utilizada, existindo estudos que utilizaram desde o equivalente a 0,1 (Figueiredo et al., 2008) até 11 amostras por hectare (Carvalho et al., 2009). Assim sendo, o objetivo desse trabalho foi determinar o efeito da densidade de amostragem de dados em mapas no treinamento de modelos preditores sobre a acurácia dos modelos de AD e sobre o número de unidades de mapeamento de solo (UMS) preditas na geração de mapas digitais de tipos de solos em duas bacias hidrográficas no noroeste do Estado do Rio Grande do Sul.

4.2 Materiais e Métodos

As áreas utilizadas para a realização desse estudo foram as bacias hidrográficas do Rio Santo Cristo e do Rio Lageado Grande, cujas características e distribuição de solos foram apresentadas no material e métodos (ítems 3.1.1 e 3.1.2) e as variáveis utilizadas também são as mesmas, sete, que foram descritas no materiais e métodos (item 3.2) e com as mesmas características.

O objetivo principal desse estudo foi testar a resposta dos modelos frente a diferentes densidades de amostragens, entretanto, juntamente com o teste da densidade de amostragens foram testados outras interações com o conjunto de dados para avaliar se os modelos se comportariam de forma semelhante ou não. Essas outras interações foram baseadas em relação à área para a realização dos cálculos das variáveis e com relação às grandezas das variáveis. Com relação às grandezas das variáveis gerou-se modelos com as variáveis com suas grandezas originais e posteriormente com os dados

normalizados e, com relação a área de cálculos das variáveis, foram utilizados variáveis com os valores calculados somente dentro da área da bacia e com valores calculados com uma área além da área da bacia. A explicação de como foi realizados esses procedimentos será dada adiante.

A amostragem dos dados a serem utilizados no treinamento do modelo de predição consistiu de coleta de informações sobre os mapas gerados, tanto sobre os mapas das variáveis preditoras como sobre o mapa da variável resposta, que era o mapa de solos convencional. Assim, os pontos de amostragem consistiram de posições geográficas exatas onde foram coletadas todas as informações sobre as variáveis ambientais. A amostragem foi feita utilizando-se diferentes densidades amostrais com pontos distribuídos de forma aleatória utilizando-se a função *random* do ArcGis 9.3 na proporções no mapa equivalentes a 0,1; 0,3; 1; 1,5; 2; 3 e 4 pontos por hectare no terreno. Essas densidades amostrais correspondem a 9, 30, 90, 135, 180, 270 e 360 mil pontos na bacia do Rio Santo Cristo e a 5, 15, 50, 75, 100, 150 e 200 mil pontos na bacia do Rio Lageado Grande.

Para a avaliação da influencia da resposta dos modelos na área utilizada para a realização do cálculo das variáveis foram utilizados dois procedimentos. Num primeiro momento realizou-se o cálculo utilizando-se somente o MDS na área interna da bacia. Para tanto utilizou-se a máscara da área da bacia para o recorte do MDS. Posteriormente, realizou-se os cálculos das variáveis utilizando-se uma área maior do que a área interna da bacia que, para isso, foi desenhado uma máscara maior que a área da bacia para a realização do recorte do MDS. Para a realização dos recortes foi utilizado a ferramenta *Extract by Mask* do ArcGis 9.3. Após a realização desses cálculos procedeu-se a normalização dos mesmos. Para a tanto, utilizou-se a média e o desvio padrão dos mapas de cada variável aplicando-se a seguinte fórmula:

$$Vn = \frac{(vo - \mu)}{\sigma}$$

Onde Vn é a variável normalizada, vo a variável original, μ a média da variável e σ é o desvio padrão. Esse procedimento foi realizado, pois cada uma das variáveis calculadas apresentam escalas de valores e unidades de medidas diferentes. Dessa maneira todas as variáveis ficaram com média 0

(zero) e desvio padrão 1 (um) e assim, com todas as variáveis na mesma grandeza pode-se avaliar se os modelos respondem de maneira diferente a cada conjunto de variável. Para a realização desse procedimento foi utilizado a ferramenta Raster Calculator do ArcGis 9.3.

A amostragem propriamente dita foi realizada gerando-se o número de pontos aleatórios necessários para cada densidade amostral avaliada e, nessas localizações espaciais, os dados sobre cada uma das variáveis foram coletados e tabulados utilizando-se a função *Sample* do ArcGis 9.3. Os dados tabulados consistiram de tabelas nas quais cada ponto amostral constituiu-se numa linha de tabela e cada coluna correspondia a uma variável amostrada. Essas tabelas foram exportadas para proceder ao treinamento dos modelos de predição com o algoritmo Simple Cart no programa Weka 3.6.3 (Hall et al., 2009). A escolha do algoritmo Simple Cart deveu-se ao fato que, dos algoritmos testados (J48 e Simple Cart) para as bacias em análise, foi o que resultou nos melhores valores de acurácia geral, concomitantemente com a geração de menores ADs. Foram testados 7 valores de número mínimo de elementos no nó final (nas ADs no Weka, sendo eles: 2, 10, 50, 100, 200, 300 e 400). A variação do M intencionou avaliar o efeito da complexidade das ADs na acurácia dos modelos preditores, dado que valores de M maiores geram ADs menores, que são mais fáceis de implementar no ArcGis para a produção dos mapas quando esse processo é feito de forma manual. O teste utilizado para a seleção do modelo foi o *Cross Validation* (10 folds) e foi calculada a acurácia geral (AG) a partir da matriz de erros de Congalton (1991), que representa a concordância entre o mapa de referência e a aplicação do modelo na geração de um mapa digital de solos

4.3 Resultados e discussão

A aplicação dos sete valores de elemento no nó final com sete diferentes densidades de pontos por hectare com mais quatro diferentes formas de obtenção de dados (variáveis normalizadas e não normalizadas nas áreas interna ou externa às bacias). Entretanto, como não foram observadas diferenças entre os resultados obtidos entre as diferentes formas de obtenção dos dados serão apresentados somente os resultados referentes ao dados

normalizados dentro das áreas nas duas bacias. Os demais dados estão apresentados nos Apêndices 1, 2, 3, 4, 5 e 6.

Na Figura 7 (a-b-c) e na Figura 8 (a-b-c) são apresentados os resultados referentes ao tamanho das ADs, às acurácias gerais (AG) e ao número de UMS preditas nas diferentes densidades de pontos amostrais nas Bacias do Santo Cristo e Lageado Grande, respectivamente.

Em relação ao tamanho das ADs da bacia do Rio Santo Cristo (Figura 7a) observou-se que com o aumento da densidade amostral houve aumento no tamanho das árvores com todos os valores de M testados, sendo geradas desde árvores com menos de 200 folhas, na densidade de 0,1 pontos por hectare, até árvores com mais de 25.000 folhas, na densidade de 4 pontos por hectare. Quando analisamos os resultados dos tamanhos das ADs na bacia do Rio Lageado Grande (Figura 8a) observou-se que, de modo semelhante à bacia do Rio Santo Cristo, ocorreu um aumento no número de folhas em função do aumento da densidade amostral, apresentando ADs com menos de 100 folhas na densidade de 0,1 pontos por hectare para mais de 14.000 folhas na densidade de 4 pontos por hectare. O aumento no tamanho das árvores com o aumento da densidade amostral também foi observado por Moran & Bui (2002) e por ten Caten et al., (2013), que destacam que esses modelos de AD possuem uma grande capacidade preditiva, porém com uma grande complexidade.

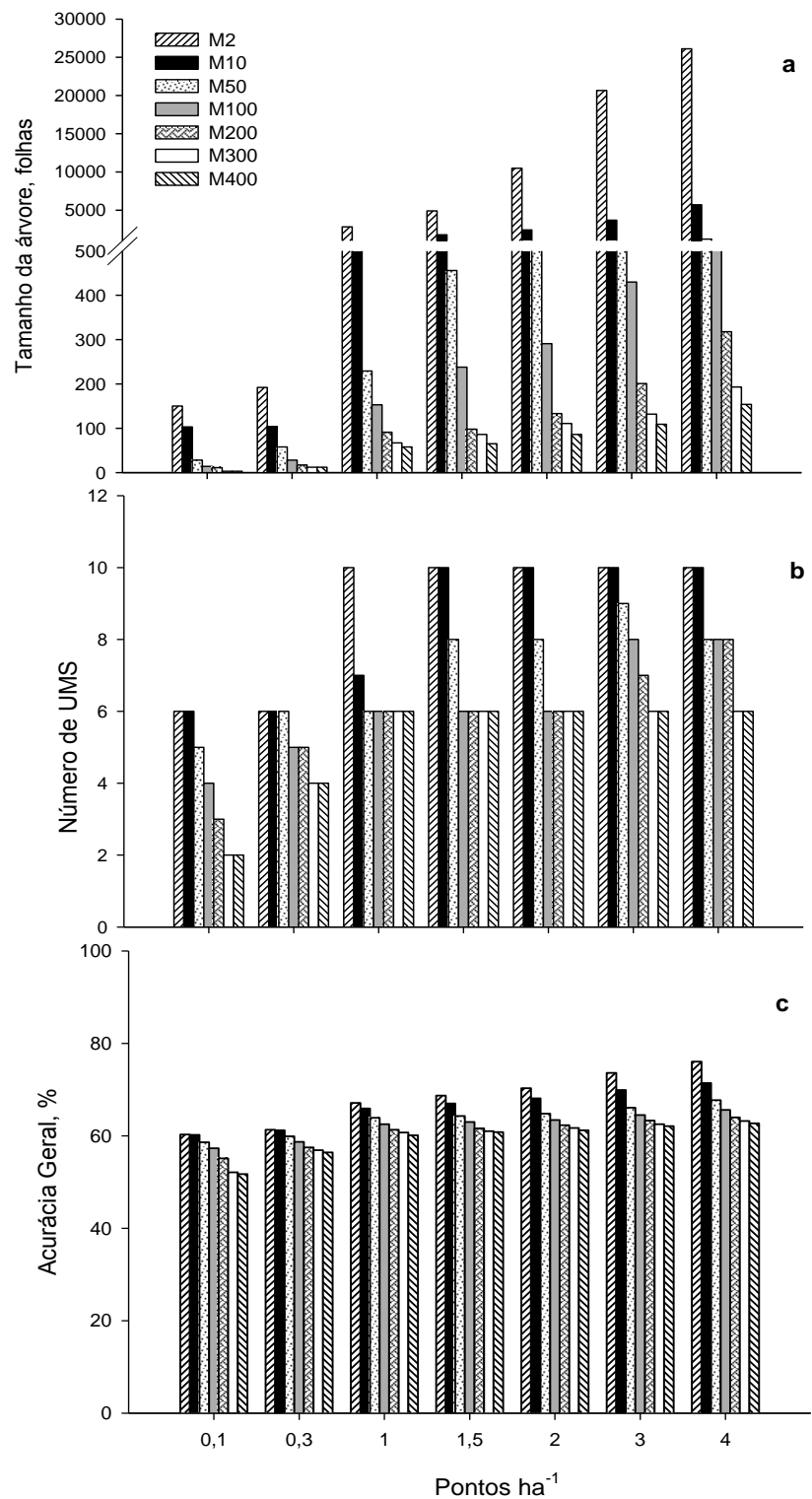


FIGURA 7. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos (UMS) previstas (b) e acurácia geral nas diferentes densidades de amostragens na Bacia do Santo Cristo.

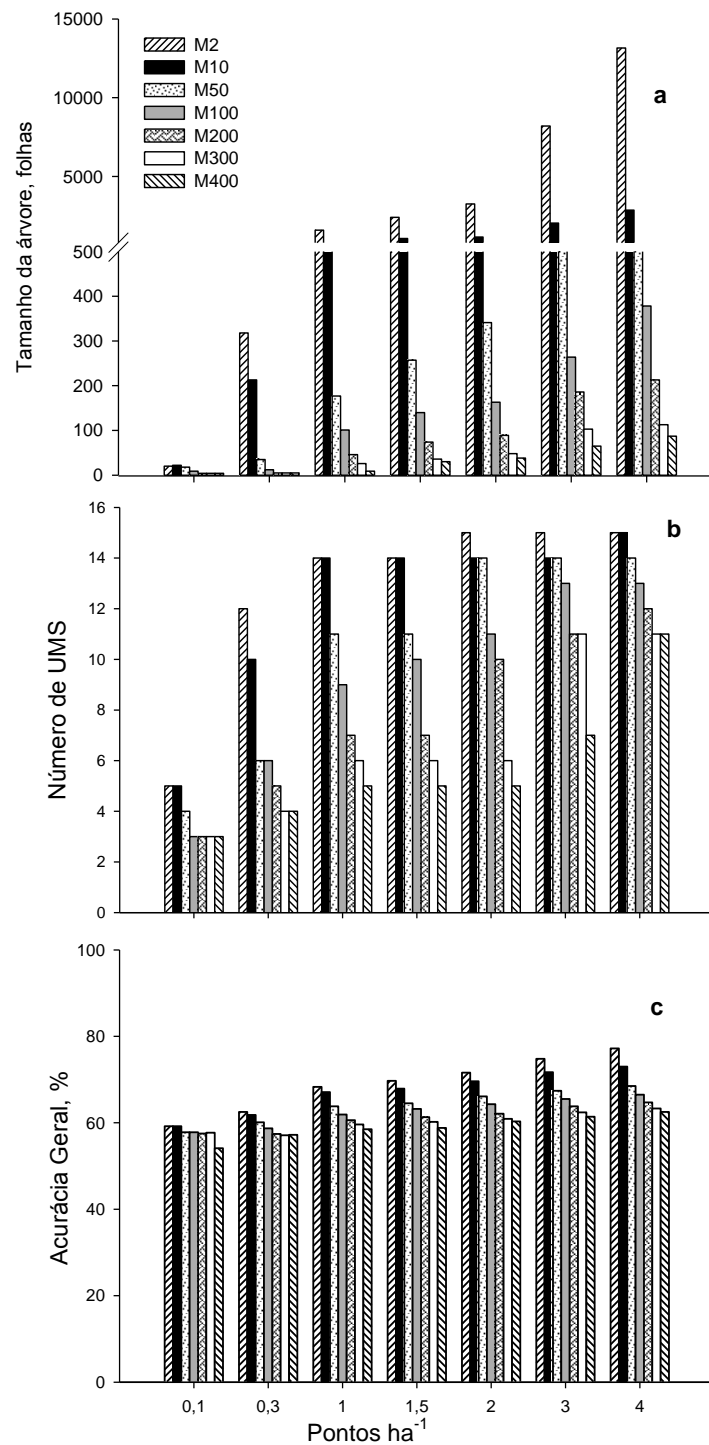


FIGURA 8. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos preditas (UMS) (b) e acurácia geral (c) nas diferentes densidades de amostragens na Bacia do Lageado Grande.

Quando se analisam os dados referentes ao número de UMs preditas na bacia do Rio Santo Cristo (Figura 7b) por cada um desses modelos gerados observa-se que, a partir de 1 ponto por hectare, com $M=2$, a totalidade das UMs (10) são preditas. Nas densidades menores que 1 ponto por hectare os modelos conseguiram predizer somente seis das dez UMs (G, LV1, LV2, RR1, RR2 e RR3). Todavia, vale salientar que a área total ocupada por essas seis UMs equivale a 99,3 % da área da bacia (Tabela 3) e que, portanto, as demais UMs provavelmente não foram preditas devido à pequena extensão das mesmas na bacia. Em relação ao número de UMs preditas na bacia do Rio Lageado Grande (Figura 8b), observa-se que somente a partir de 2 pontos por hectare ocorreu a predição de todas as 15 UMs e somente nas ADs mais complexas obtidas com $M=2$. Observou-se também que na menor densidade de amostragem (0,1 pontos por hectare) o modelo conseguiu prever somente cinco UMs (G, LV1, LV2, RR1 e RR2) o que pode ter ocorrido porque as 10 UMs não preditas ocupam somente 7,5 % da área (Tabela 3). Assim sendo, quando da utilização de baixa densidade de pontos amostrais, o modelo provavelmente não conseguiu captar a variabilidade presente, conforme também observado por Moran e Bui (2002).

Em relação à AG, na bacia do Rio Santo Cristo (Figura 7c) observou-se que a mesma variou de 60 a 76 % da menor para a maior densidade de pontos amostrais. Observou-se também que a maior diferença (de 6 %) ocorreu entre as densidades de 0,3 pontos por ha e de 1 ponto por hectare, enquanto que o incremento entre as demais densidades foi de 2 %. Quando se observa a AG na bacia do Rio Lageado Grande (Figura 8c), observa-se que a mesma variou de 54 %, na menor densidade, a 74 % na maior densidade de amostragem. O comportamento de incremento da AG foi semelhante ao observado na bacia do Santo Cristo.

Esse aumento da acurácia com o aumento do tamanho do conjunto de dados para treinamento também foi observado por Pal e Mather (2002) e por Grinand et al., (2008). Esses últimos autores também observaram que o maior incremento se deu entre as amostragens com volumes de dados menores (entre 10 a 20 %), sendo que com volumes de dados maiores os autores não observaram aumentos significativo na qualidade dos modelos. Resultados semelhantes a esses também foram observados por ten Caten et al., (2013).

De um modo geral observou-se que o aumento do tamanho das ADs, principalmente com $M=2$, em todas as densidades, não refletiu no aumento proporcional das acurácias em ambas as bacias. Por outro lado, quando se observou os dados referentes à ADs menos complexas com tamanhos entre 150 e 200 folhas, na bacia do Rio Santo Cristo, não se observaram grandes diferenças nos valores de acurácia (60 a 63 %) e nem diferenças no número de UMS preditas (seis) entre as diferentes densidades de amostragem. Já na bacia do Rio Lageado Grande notou-se que, igualmente, nas ADs menos complexas, não houve aumento na acurácia geral como aumento da densidade de amostragem, porém ocorreu um aumento do número de UM preditas. Na menor densidade de amostragem (0,1 pontos por hectare) foram preditas somente cinco UMs (G, LV1, LV2, RR1 e RR2), enquanto que, nas maiores foram preditas todas as 15 UMs.

Assim sendo, percebe-se que, nas duas bacias, o aumento da densidade de amostragem influenciou a AG e o número de UMs preditas. Nas Ads menos complexas geradas, percebe-se que quando há uma dominância de classes (como na bacia do Rio Santo Cristo) a AG e o número de UMS preditas não foram influenciadas pela densidade de pontos. Entretanto, quando existe maior número de UMs de menores extensões (como na bacia do Rio Lageado Grande), o aumento da densidade de pontos proporcionou um aumento no número de UMs preditas, porém sem aumento da acurácia. O aumento de UMs preditas provavelmente ocorreu porque, com o aumento do número de pontos de amostragem, uma maior quantidade deles foi alocada nas UMs com extensões intermediárias e, com esse aumento, o modelo conseguiu prever mais UMs.

Portanto, analisando os dados das duas bacias, observa-se que, a utilização da densidade de amostras de pelo menos 1 ponto por hectare é o suficiente para captar a variabilidade nas áreas. Por outro lado, se houver a necessidade de controlar o tamanho das AD, é indiferente utilizar qualquer uma das densidades de amostragem para obter valores de acurácia geral maiores que 60 %.

4.4 Conclusões

Quando não foi limitado o tamanho das árvores de decisão, o uso de densidades de amostragem maiores resultou no aumento da concordância dos mapas digitais preditores de tipos de solos gerados com os mapas convencionais de solos usados como referência e no aumento do número de unidades de mapeamento de solos preditas.

Quando limitou-se a complexidade das árvores de decisão para facilitar sua implementação manual para a geração dos mapas em SIG, a variação nas densidades de amostragem não influenciou a capacidade de gerar-se mapas concordantes com os mapas de solos de referência e influenciou muito pouco o número de unidades de mapeamento de solos preditas.

O uso de densidades de amostragem de menos de uma amostra por hectare demonstrou ser inadequado para a geração de mapas digitais de solos, principalmente quando nos mapas de referência ocorrer maior número de UMs.

O tamanho das árvores de decisão foram afetadas pelo tamanho do conjunto de dados e pelo número de elementos no nó final, sendo que, quanto maior o tamanho do conjunto de dados e menor o número de elementos no nó final maior a árvore de decisão.

5 CAPÍTULO II – RESPOSTA DOS MODELOS DE ÁRVORE DE DECISÃO COM DIFERENTES CONJUNTOS DE VARIÁVEIS UTILIZADAS EM MAPEAMENTO DIGITAL DE SOLOS

5.1 Introdução

Os estudos em mapeamento digital de solos, assim como os levantamentos pedológicos convencionais, se baseiam nas relações solo-paisagem. Porém, enquanto nos levantamentos convencionais as relações solo-paisagem são estabelecidas de forma qualitativa, no mapeamento digital elas são avaliadas quantitativamente a partir de variáveis geomorfométricas derivadas de Modelo Digital de Elevação (MDE) (Moore et al., 1991). Assim sendo, o mapeamento digital de solos (MDS) se baseia na geração de sistemas de informações que permitem estabelecer relações matemáticas entre atributos do terreno e classes de solos e, assim, prever a distribuição espacial das classes de solos com bastante semelhança aos mapas gerados por métodos convencionais (Lagacherie e McBratney, 2007).

As variáveis geomorfométricas derivadas do MDE podem ser diferenciadas entre variáveis primárias e secundárias (Moore et al., 1993): as variáveis primárias são aquelas derivadas diretamente do MDE, sendo elas: elevação, declividade, orientação, curvatura de perfil, curvatura planar, comprimento do fluxo e acúmulo de fluxo. As variáveis secundárias são as que envolvem a combinação de variáveis primárias, geralmente índices geomorfométricos, tal como o índice de umidade topográfica (IUT) (Beven & Kirkby, 1979) e *Stream Power Index* (Moore et al., 1988), entre outros.

Esses atributos do terreno são amplamente utilizados como preditores no MDS (McBratney et al., 2003), pois o relevo é um dos fatores pedogenéticos que tem grande influência na formação dos solos (Jenny 1941) e também porque os MDE estão disponíveis para praticamente todas as regiões, com diferentes resoluções e de forma gratuita (Behrens et al., 2010). Apesar da

disponibilidade desse tipo de dados, poucos estudos foram realizados para se saber qual atributo do terreno é mais importante na predição correta das classes de solos em diferentes tipos de relevo e, se existe diferença entre as acurácias dos modelos, quando da utilização de diferentes variáveis. A seleção e a combinação das variáveis a serem utilizadas para a construção de modelos de predição em MDS pode ser um trabalho difícil (Lark et al., 2007), pois para cada conjunto diferente de dados são gerados modelos com diferentes acurácias (Sun et al., 2011).

Vários métodos de modelagem são utilizados no MDS (McBratney et al., 2003). Entre eles, as árvores de decisão são frequentemente utilizadas para a previsão de classes de solo (Henderson et al., 2005; Scull et al., 2005; Grinand et al., 2008). A alimentação do modelo com um adequado conjunto de variáveis é fundamental pois os preditores utilizados devem apresentar uma relação plausível com o solo, com base no conhecimento do sistema solo-paisagem (Kempen et al., 2009) e também porque cada variável pode ser mais importante para um determinado ambiente, enquanto em outros a importância da mesma pode ser menor (Behrens et al., 2010). Com o intuito de verificar a influência de um conjunto de variáveis na resposta do algoritmo de árvore de decisão numa região montanhosa de Hong Kong, Sun et al. (2011) realizaram uma combinação entre as variáveis e identificaram que as variáveis que mais influenciaram o resultados nos modelos de árvore de decisão foram material de origem, elevação, declividade, curvatura do perfil e área de contribuição da bacia. Behrens et al. (2010), utilizando um conjunto de 18 variáveis constataram que a utilização de somente duas das variáveis (perfil relativo da curvatura e índice topográfico combinado) foi suficiente para a predição de ocorrência de solos fluviais. As demais classes de solo, entretanto, foram preditas por uma mistura das variáveis. Além disso, constataram também que a variável declividade teve uma maior importância na predição de solos que ocorriam em áreas montanhosas. Ainda, como estudos dessa natureza são escassos ainda são necessários estudos para se analisar o comportamento dos modelos perante um conjunto de variáveis em áreas com características de relevo e distribuição de classes de solos diferentes. Assim sendo, o objetivo deste trabalho foi identificar quais as variáveis que mais influenciam a resposta

dos modelos de árvore de decisão em duas bacias hidrográficas do Rio Grande do Sul que apresentam características ambientais distintas.

5.2 Materiais e métodos

Para a realização desse estudo foram utilizadas as bacias do Santo Cristo e do Arroio Portão. Utilizou-se essas bacias pois as mesmas apresentam diferenças no que se refere à área da bacia, ao relevo, ao material de origem, ao clima, e, conseqüentemente, aos solos. As características das bacias encontram-se descritas no item 3.1.1 e item 3.1.3.

Para a geração dos modelos foram testadas 11 variáveis: elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

As amostras para treinamento dos algoritmos foram geradas de maneira aleatória e corresponderam a densidade de três pontos amostrais por hectare para ambas as bacias. Na bacia do Santo Cristo foram amostrados 270.000 pontos e na bacia do Arroio Portão foram amostrados 60.000 pontos, nos quais foram registrados valores de todas as variáveis preditoras citadas acima e da unidade de mapeamento de solo (UMS) ocorrente. A escolha dessa densidade de amostragem deu-se em função dos resultados do estudo 1.

Os dados foram exportados para proceder ao treinamento dos modelos com o algoritmo *Simple Cart* no programa Weka 3.6.3 (HALL et al., 2009). Foram gerados os modelos com número de elementos no nó final (*minNumObj* - (M) igual a 2, sendo que, quando M é igual a 2 ocorre a geração de árvores de decisão de maiores complexidades e maiores acurácias. Os modelos foram gerados adotando-se três procedimentos diferentes. Primeiramente gerou-se modelos com cada variável individualmente, a fim de se detectar qual variável, por si só, tem o maior poder preditivo. Em seguida, das 11 variáveis rodou-se modelos retirando-se uma delas para se detectar qual tem a maior importância no conjunto e, por último, realizou-se combinações de duas em duas variáveis. Nesse processo foram gerados 76

diferentes modelos para cada bacia, nos quais foram avaliados o tamanho das ADs geradas, o número de UMS preditas e a acurácia global (AG). Após a avaliação da resposta dos modelos gerou-se ADs com as variáveis que apresentaram os melhores resultados com a intenção de avaliar a capacidade de resposta desse conjunto de variáveis.

5.3 Resultados e discussão

Nas Tabelas 5 e 6 são apresentados os resultados do tamanho das ADs, acurácia geral (AG), quantidade de UMS preditas e listagem de UMS preditas referentes à bacia do Santo Cristo e Arroio Portão, respectivamente, quando os modelos foram gerados com somente uma das variáveis.

Analisando os resultados (Tabelas 5 e 6) pode-se observar que o tamanho das ADs variaram em função da variável utilizada, entretanto, na maioria dos casos, foram menores do que a AD gerada com a totalidade das variáveis. Os valores de AG apresentaram uma diminuição de aproximadamente 30% e, da mesma maneira pode-se observar que a quantidade de UMS preditas também diminuiu. Apesar dos valores de AG serem menores percebe-se que algumas variáveis tiveram maior capacidade de predição, ou seja, maior valor de AG em relação às outras. Dentre essas destaca-se o comprimento de fluxo (CF), a elevação do terreno (MNT) e a distância de rios (DIST) tanto na bacia do Santo Cristo como na do Arroio Portão. Entretanto, apesar dos modelos gerados com essas variáveis apresentarem valores de acurácia maiores, percebe-se que na bacia do Santo Cristo, somente o CF conseguiu predizer a maioria das UMS presentes na bacia, as demais conseguiram separar somente as UMS com maior representatividade, ou seja, os Latossolos (LV1) dos Neossolos (RR1). As UMS que não foram preditas com a utilização do CF foram as UMS M1, RL e RR4. Essas UMS são pouco representativas na bacia, sendo que juntas correspondem a menos de 0,5% da área total da mesma. Essa dificuldade dos modelos predizerem as UMS com baixa representatividade também foi constatado por Sarmiento (2010) em estudo realizado no Vale dos Vinhedos em Bento Gonçalves/RS. Neste sentido, Bailey et al. (2003) destacam para uma melhor predição dos modelos foi necessário a eliminação das classes que ocupavam uma área menor que 5% da área total do mapa original.

Na bacia do Arroio Portão os modelos gerados pelas variáveis CF e o MNT conseguiram prever seis das nove UMS. As UMS não previstas nesta bacia foram a CHa, SXd e Aluvial que, assim como na bacia Santo Cristo, são as UMS menos representativas, juntas correspondendo a 1,3% da área total. Contudo, mesmo que essas variáveis tenham conseguido prever quase a totalidade das UMS, as AGs desses modelos não ultrapassaram 53% na bacia do Santo Cristo (Tabela 5) e de 54% na bacia do Arroio Portão (Tabela 6). Estes valores de AG são, aproximadamente, vinte pontos percentuais mais baixo quando comparadas com as ADs com todas as variáveis, que foram de 72,6% e 71,2% respectivamente na bacia do Santo Cristo e na bacia do Arroio Portão.

Assim sendo, a geração dos modelos indicaram que existem algumas variáveis com maior capacidade de predição do que outras no conjunto avaliado, entretanto, essas mesmas variáveis, individualmente, não apresentam capacidade preditiva igual ao conjunto com todas as variáveis.

A retirada de uma das variáveis do conjunto de dados praticamente não afetou a resposta dos modelos (Tabelas 7 e 8) tanto no que tange ao tamanho das ADs, nos valores de AG e na quantidade de UMS previstas, em ambas bacias. A exceção a esses resultados ocorreu quando as variáveis CF, MNT e DIST foram retiradas dos modelos tanto na bacia do Santo Cristo como na do Arroio Portão. A retirada das mesmas acarretou a diminuição da AG de 72,6% para 57,5%; 61,8%; 69,4%, na bacia do Santo Cristo, e de 71,2% para 59,9%; 61,1% e 66,9%, na bacia do Arroio Portão. Além da diminuição dos valores de AG, a retirada dessas variáveis também ocasionou a diminuição do número de UMS previstas. As UMS que não foram previstas, igualmente aos resultados obtidos anteriormente, foram aquelas que apresentam menor área, ou seja, as UMS, RL, M1, RR4 e RR5, na bacia do Santo Cristo e CHa, SXd e Aluvial, na bacia do Arroio Portão.

Resultados semelhantes a esses também foram observados por Sun et al. (2011). Esses autores observaram que a retirada das variáveis menos importantes não acarretou em diferenças na acurácia e, também, ocorreu pouca mudança no tamanho das ADs. Entretanto, quando realizaram a retirada das variáveis mais importantes os valores de acurácia e os tamanhos das ADs reduziram-se bastante.

Das variáveis que mais influenciaram as respostas dos modelos a relação da altitude na formação dos solos pode ser entendida através do conceito de catena proposta por Milne (1935), a qual, pode ocorrer a formação de diferentes solos dependendo da sua posição na paisagem, devido a movimentação da matéria devido ao movimento e distribuição de água na mesma, na qual podem ocorrer áreas de erosão, nos locais mais íngremes e áreas de acúmulo nos locais mais planos (Kampf e Curi, 2012), acarretando assim a formação de diferentes tipos de solos. A distância dos rios também tem grande influência quanto a posição do lençol freático, que normalmente está na superfície ou próxima dela na proximidade dos rios, enquanto que, com a distância dos mesmos ou nas regiões mais altas o lençol freático torna-se mais baixo, ou seja, o perfil do solo permanece acima do lençol freático e, em consequência disso ocorre a presença de solos mais bem drenados (Kampf e Curi, 2012).

TABELA 5. Resultados do tamanho das árvores de decisão, número de UMS, acurácia geral (AG) e UMS previstas na bacia do Santo Cristo dos modelos gerados com somente um das variáveis.

Variáveis											Tamanho da árvore	AG	nº UMS	UMS previstas
CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	SPC	LC	TPI				
											10545	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
											17	39	2	LV1;RR1
											136	50	2	LV1;RR1
											4	41	2	LV1;RR1
											17627	49	7	G:LV2;LV1;RR2;RR3;RR1;RR5
											181	53	4	LV2;LV1;RR2;RR1
											127	43	3	LV1;RR2;RR1
											2643	43	6	G:LV2;LV1;RR2;RR3;RR1
											344	42	3	LV1;RR2;RR1
											7	40	2	LV1;RR1
											9	43	2	LV1;RR1
											143	45	2	LV1;RR1

*elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

TABELA 6. Resultados do tamanho das árvores de decisão, número de UMS, acurácia geral (AG) e UMS preditas na bacia do Arroio Portão dos modelos gerados com somente um das variáveis.

Variáveis											Tamanho da árvore	AG, %	nº UMS	UMS preditas
CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	SPC	LC	TPI				
											1198	71,2	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe; CHa; SXd; Aluvial
											45	41,2	3	PVa; PVd; PVa/GMe
											64	44,5	4	PVa; PVd; GMd; PVa/GMe
											39	41,4	2	GMd; PVa/GMe
											38	50	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											128	53,9	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											6	42,9	3	PVd; CXf/RLe; PVa/GMe
											10	41,7	3	PVd; CXf/RLe; PVa/GMe
											123	41,9	3	PVd; PVa; PVa/GMe
											6	41,3	2	PVa; PVa/GMe
											11	41,8	3	PVd; PVa; PVa/GMe
											303	45,6	5	PVd; PVa; PVa/GMe; Mxo/RLe; GMe

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC)

TABELA 7. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo retirando uma das variáveis do conjunto de dados

Combinação	Variáveis											Tamanho da árvore	AG, %	nº UMS	UMS previstas
	CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	LC	SPC	TPI				
												10545	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10511	72,7	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												11061	69,4	9	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5
												10473	72,9	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												577	57,5	6	G:LV2;LV1;RR2;RR3;RR1
												6504	61,8	8	G:LV2;LV1;M1;RR2;RR3;RR1;RR5
												10682	72,5	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10482	72,3	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10626	70,2	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10538	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10544	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												10675	72,1	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

TABELA 8. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Arroio Portão retirando uma das variáveis do conjunto de dados.

Combinação	Variáveis										Tamanho da árvore	AG, %	nº UMS	UMS previstas	
	CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	LC	SPC					TPI
												1198	71,2	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1332	71,2	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1217	66,9	8	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd
												1178	71,3	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												249	59,9	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												1230	61,1	7	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe; SXd
												1353	71,1	8	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd
												1229	71,4	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1261	70,8	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1196	71	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1198	71,3	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial
												1315	71,5	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa; SXd; Aluvial

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

Em relação às combinações duas a duas (Tabelas 9 e 10) observa-se que nenhuma das combinações geraram modelos com a capacidade de prever a totalidade das UMS. Entretanto, algumas combinações geraram modelos com maior capacidade preditiva do que outras, sem contudo, gerar modelos com AG que se equivalassem aos resultados do modelo com a totalidade das variáveis. A maioria dos valores de AG não ultrapassaram 50%.

As combinações também não conseguiram gerar modelos que conseguissem prever a totalidade das UMS (Tabelas 9 e 10). Algumas combinações conseguiram prever no máximo seis UMS na bacia do Santo Cristo (Tabela 9) e oito na bacia do Arroio Portão (Tabela 10). A grande maioria das combinações conseguiram prever duas ou três UMS, sendo essas as mais representativas nas bacias. Dentre os modelos que geraram os melhores resultados, tanto com relação aos valores de AG como na quantidade de UMS previstas, observa-se que as variáveis comprimento de fluxo e elevação do terreno estiveram presentes nas combinações. Os melhores resultados, entretanto, foram obtidos na combinação dessas duas variáveis, tanto na bacia do Santo Cristo como na bacia do Arroio Portão. Nessa combinação, a AG foi de aproximadamente 65%, valor esse muito próximo ao valor com todas as variáveis.

Após a análise de todas as combinações selecionou-se as variáveis que mais influenciaram nas resposta dos modelos e gerou-se um modelo a partir delas. Para tal foram utilizadas as variáveis CF, MNT e DIST nas duas bacias estudadas. Os resultados mostram que tanto na bacia do Santo Cristo como na bacia do Arroio Portão (Tabela 11), os valores de AG foram muito semelhantes com os valores de AG com o conjunto todo. Comportamento semelhante foi observado por ten Caten et al. (2011b) quando realizaram a diminuição de variáveis no conjunto. A diminuição do número de variáveis utilizadas para a construção dos modelos facilita a compreensão das ADs, pois forma modelos com menos complexidade e gera os modelos com maior rapidez, já que, existe um menor número de variáveis a serem analisadas.

Em estudo realizado por Thomas et al., (1999), com a finalidade de avaliar a capacidade de predição de solos com um conjunto de variáveis geomorfométricas em duas bacias na França, também foi evidenciado que algumas variáveis são mais importantes que outras dentro de um conjunto.

Neste estudo foi observado que em uma das bacias as variáveis aspecto e declividade foram as que mais contribuíram para explicar a distribuição espacial dos solos e que na outra bacia, em contrapartida, foram as variáveis curvatura do perfil e distância dos rios. Behrens et al. (2010), avaliando um conjunto de 18 variáveis observaram que os solos fluviais foram bem preditos utilizando somente duas variáveis (índice topográfico composto e perfil relativo da curvatura), enquanto que as demais classes de solos foram preditas por uma combinação de várias variáveis. A declividade se mostrou uma variável importante na predição de classes de solos que ocorrem em grandes declividades.

TABELA 9. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo com a combinação de variáveis duas a duas para geração do modelo de árvore de decisão.

	Variáveis										tamanho da árvore	AG	nº UMS	UMS previstas	
	CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	IC	SPC					TPI
Combinções												10545	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												363	52,5	6	G:LV2;LV1;RR2;RR3;RR1
												431	56,4	6	G:LV2;LV1;RR2;RR3;RR1
												3829	64,8	6	G:LV2;LV1;RR2;RR3;RR1
												571	50,5	6	G:LV2;LV1;RR2;RR3;RR1
												150	53,1	6	G:LV2;LV1;RR2;RR3;RR1
												349	53,6	6	G:LV2;LV1;RR2;RR3;RR1
												1431	53,3	6	G:LV2;LV1;RR2;RR3;RR1
												307	51,4	6	G:LV2;LV1;RR2;RR3;RR1
												304	50,4	6	G:LV2;LV1;RR2;RR3;RR1
												129	49,2	5	LV2;LV1;RR2;RR3;RR1
												331	52,6	5	LV2;LV1;RR2;RR3;RR1
												1663	41,9	5	LV2;LV1;RR2;RR3;RR1
												326	52,8	5	G:LV2;LV1;RR2;RR1
												1483	52,8	5	LV2;LV1;RR2;RR3;RR1
												208	50,0	5	LV2;LV1;RR2;RR3;RR1
												38	52,8	5	LV2;LV1;RR2;RR3;RR1
												70	55,4	4	LV2;LV1;RR2;RR1
											49	53,8	4	LV1;LV2;RR1;RR2	

Continuação.....

Combinções												1397	52,7	4	LV2;LV1;RR2;RR1
												125	49,7	3	RR1;RR2;LV1
												136	41,4	3	RR1;RR2;LV1
												65	42,7	3	RR1;RR2;LV1
												777	43,9	3	RR1;RR2;LV1
												152	50,8	3	LV1;RR1;RR2
												440	43,7	3	LV1;RR1;RR2
												46	43,7	3	LV1;RR1;RR2
												52	49,8	3	LV1;RR1;RR2
												66	43,0	3	RR1;LV1; RR2
												288	44,8	3	RR1;LV1; RR2
												401	43,0	3	RR1;LV1; RR2
												804	46,4	3	RR1;LV1; RR2
												156	49,7	2	RR1;LV1
												713	44,0	2	RR1;LV1
												460	42,0	2	RR1;LV1
												355	46,4	2	RR1;LV1
												513	42,9	2	RR1;LV1
												90	43,2	2	RR1;LV1
												276	41,9	2	RR1;LV1
											340	50,4	2	RR1;LV1	

Continuação.....

Combinções												20	43,3	2	RR1;LV1
												222	47,5	2	RR1;LV1
												548	43,5	2	RR1;LV1
												94	43,4	2	RR1;LV1
												49	43,3	2	RR1;LV1
												99	49,8	2	RR1;LV1
												149	45,8	2	RR1;LV1
												5499	42,4	2	RR1;LV1
												44	41,5	2	RR1;LV1
												39	40,3	2	RR1;LV1
												47	49,7	2	RR1;LV1
												976	46,0	2	RR1;LV1
												501	46,3	2	RR1;LV1
												241	45,1	2	RR1;LV1
												241	45,1	2	RR1;LV1
											800	51,0	2	RR1;LV1	

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

TABELA 10. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Arroio Portão combinação de variáveis duas a duas para geração do modelo de árvore de decisão.

Combinação	Variáveis										tamanho da árvore	AG	nº UMS	UMS preditas	
	CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	IC	SPC					TPI
												1198	71,2	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe; CHa; SXd; Aluvial
												904	64,9	8	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe; CHa; SXd
												67	50,3	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												115	54,3	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												26	50,4	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												196	54,1	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												763	58,9	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												73	51,5	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												212	54,3	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												67	58,0	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												38	55,0	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												20	54,2	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												504	45,9	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												406	54,6	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												45	51,0	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
												47	50,4	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe

Continuação....

Combinação											44	54,5	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											402	43,3	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											553	45,1	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											116	50,7	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											888	44,6	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											249	51,7	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											73	55,1	6	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe
											269	42,3	5	PVa; PVd; CXf/RLe; GMd; PVa/GMe
											53	46,8	5	PVa; PVd; CXf/RLe; GMd; PVa/GMe
											19	46,1	5	PVa; PVd; CXf/RLe; GMd; PVa/GMe
											99	43,4	5	PVa; PVd; MXo/RLe; GMd; PVa/GMe
											39	44,1	5	PVa; PVd; MXo/RLe; CXf/RLe; PVa/GMe
											25	45,6	5	PVa; PVd; GMd; CXf/RLe; PVa/GMe
											31	42,3	5	PVa; PVd; GMd; CXf/RLe; PVa/GMe
											73	43,2	5	PVa; PVd; GMd; CXf/RLe; PVa/GMe
											309	45,6	5	PVa; PVd; MXo/RLe; GMd; PVa/GMe
											707	47,0	5	PVa; PVd; MXo/RLe;; GMd; va/GMe
											138	45,2	4	PVa; PVd; GMd; PVa/GMe;
											26	43,05	4	PVa; PVd; CXf/RLe; PVa/GMe
											32	41,7	4	PVa; PVd; CXf/RLe; PVa/GMe

Continuação....

Combinação											27	45,0	4	PVa; PVd; GMd; PVa/GMe;
											126	43,2	4	PVa; PVd; CXf/RLe; PVa/GMe
											129	43,3	4	PVa; PVd; CXf/RLe; PVa/GMe
											110	42,2	4	PVa; PVd; GMd; PVa/GMe;
											122	42,0	4	PVa; PVd; MXo/RLe; PVa/GMe
											33	42,1	4	PVa; PVd; GMd; PVa/GMe
											48	41,6	4	PVa; PVd; GMd; PVa/GMe
											59	46,0	4	PVa; PVd; GMd; PVa/GMe
											127	43,3	4	PVd; PVa; CXf/RLe; PVa/GMe
											187	45,1	4	PVd; PVa; GMd PVa/GMe
											122	43,1	4	PVd; PVa; GMd PVa/GMe
											13	41,3	3	PVa; PVd; PVa/GMe;
											173	41,9	3	PVa; PVd; PVa/GMe;
											127	42,0	3	PVd; PVa/GMe;GMd
											23	41,9	3	PVd; PVa/GMe;GMd
											10	41,8	3	PVd; CXf/RLe;PVa/GMe
											6	42,6	2	PVd; PVa/GMe;
											14	41,4	2	PVa; PVa/GMe
										135	41,1	2	PVd; PVa/GMe	

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

TABELA 11. Resultados do tamanho das árvores de decisão, número de UMS e acurácia geral na bacia do Santo Cristo e do Arroio Portão com as variáveis comprimento de fluxo (CF), altitude (MNT) e distância de rios (DIST) para geração do modelo de árvore de decisão.

Combinação	Variáveis											Tamanho da árvore	AG	nº UMS	UMS preditas
	Santo Cristo														
	CF	TWI	MNT	DECL	DIST	CURV	AF	ASP	IC	SPC	TPI				
												10545	72,6	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
												12755	70,8	10	G:LV2;LV1;M1;RR2;RR3;RR1;RR4;RR5;RL
	Arroio Portão														
												1198	71,2	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa;SXd; Aluvial
												1679	71,3	9	PVa; PVd; MXo/RLe; CXf/RLe; GMd; PVa/GMe;CHa;SXd; Aluvial

* elevação do terreno (MNT), declividade (DECL), acúmulo do fluxo (AF), comprimento do fluxo (CF), curvatura (CURV), índice de umidade topográfica (TWI), distância dos rios (DIST), aspecto/orientação (ASP), classificação da posição do declive (slope position classification - SPC), índice de posição topográfica (topographic position index - TPI) e classificação das formas de relevo (landform classification - LC).

5.4 Conclusões

Os modelos gerados a partir de uma única variável apresentaram baixa acurácia e a quantidade de unidades de mapeamento de solos preditas foi pequena.

A retirada das variáveis comprimento de fluxo, elevação do terreno e distância de rios do conjunto de variáveis implicou na diminuição da acurácia, do tamanho das árvores de decisão e do número de unidades de mapeamento preditas, em ambas bacias.

Os modelos gerados a partir das variáveis mais importantes, comprimento de fluxo, elevação do terreno e distância de rios, resultaram em acurácia e quantidade de unidades de mapeamento preditas mais altos, semelhantes aos modelos gerados com todas as variáveis.

Os modelos gerados a partir das combinações duas a duas apresentaram melhores acurácias quando as variáveis comprimento de fluxo, altimetria e distância de rios estiveram presentes, entretanto em nenhuma combinação os modelos conseguiram predizer todas as unidades de mapeamento.

6. CAPÍTULO III – EXPANSÃO DE MAPEAMENTO DE SOLO PARA ÁREAS FISIOGRAFICAMENTE SEMELHANTES A PARTIR DE UM MODELO DE PREDIÇÃO TREINADO EM ÁREA DE REFERÊNCIA

6.1 Introdução

Apesar de haver uma crescente necessidade de informação de levantamentos de solos de médias a grandes escalas, os custos, os recursos humanos e o tempo necessários para a realização e execução, bem como as necessidades de recursos humanos envolvidos, tornam esses levantamentos de difícil execução. Neste sentido, devido à necessidade de diminuição de tempo e recursos financeiros para a realização desses levantamentos com esse nível de detalhamento, encontrou-se no Mapeamento Digital de Solos (MDS) uma forma para geração desses mapas através de métodos matemáticos computacionais associados a parâmetros derivados de Modelos Digital de Elevação (MDE) em sistemas de informação geográfica (Hengl et al., 2007).

Igualmente aos levantamentos pedológicos convencionais, a maior parte dos estudos em MDS baseia-se nas relações solo-paisagem. Entretanto, enquanto que nos levantamentos convencionais as relações solo-paisagem são estabelecidas de forma qualitativa, no mapeamento digital elas são avaliadas quantitativamente a partir de variáveis geomorfométricas derivadas de MDE (Lagacherie e McBratney, 2007). Assim sendo, pode-se estabelecer relações entre variáveis geomorfométricas e classes de solos por meio de modelos matemáticos, podendo-se assim prever a ocorrência de classes de solos a partir do treinamento desses modelos com o uso de mapas de solos convencionais (Lagacherie e Mcbratney, 2007), desta forma permitindo a

expansão por extrapolação do mapeamento de solos para áreas fisiograficamente semelhantes (Grinand et al., 2008).

Dentre as metodologias utilizadas para a quantificação dessas relações entre solos e variáveis ambientais, pode-se citar: a *lógica difusa* (Zhu et al., 1996), as *redes neurais artificiais* (Chagas et al., 2010), as *árvores de decisão* (Moran & Bui, 2002; Grinand et al., 2008; Giasson et al., 2011), a *geoestatística* (MacBratney et al., 2003), as *regressões logísticas* (Bailey et al., 2003; Giasson et al., 2008), entre outras. Dentre essas, apesar de ser uma técnica estatística desenvolvida somente há cerca de 30 anos (Breiman et al., 1984), as árvores de decisão estão cada vez sendo mais usadas (Lawrence et al., 2004). As principais características que tornam essa técnica atraente para uso no MDS é que são geralmente simples de entender e fáceis de interpretar e discutir, pois seguem a lógica usualmente empregada em classificações sistemáticas (Miller & Franklin, 2001; Xu et al., 2005) e são capazes de processar grandes volumes de dados praticamente sem a necessidade de pré-processamento ou transformação. Portanto, esta capacidade de processar grandes volumes de dados sem interferência humana lhe confere a capacidade de recuperar conhecimento pedológico a partir de mapas existentes, tornando-as uma eficiente ferramenta de mineração de dados (Henderson et al., 2005).

Uma vez que os modelos de AD são treinados estabelecendo-se relações entre ocorrência de solos e as variáveis ambientais, os modelos podem ser usados para realizar a extrapolação das relações solo-paisagem de uma área mapeada para áreas semelhantes ainda não mapeadas. Entretanto, até o momento, a maioria dos estudos realizados em MDS, tem buscado avaliar a capacidade dos modelos reproduzirem os mapas de solos existentes e não a sua capacidade de extrapolação. Referente a isso Grinand et al. (2008) realizaram um estudo e constataram que a capacidade preditiva do modelo utilizado foi bastante baixa para a área de validação do modelo preditor. No Brasil o único estudo dessa natureza foi realizado por ten Caten et al. (2011a) no município de São Pedro do Sul (RS), no qual constataram que a acurácia do modelo também foi menor na área de validação em comparação a área de treinamento e que, dentre as classes de solos, os melhores resultados estiveram relacionados às áreas mais declivosas onde ocorria a predominância de Neossolos. Todavia, Höfig et al. (2014), através de avaliação da verdade de

campo, encontrado acurácia do mapa predito semelhante à acurácia do mapa convencional de solos (75%) e concordância do mapa predito com o mapa convencional de 47%. Em função de haver poucos estudos relacionados a esse tema, o objetivo deste trabalho foi realizar a expansão por extrapolação do mapeamento de solos para áreas fisiograficamente semelhantes, em locais diferentes dos estudos existentes e com diferentes características de solos, utilizando-se o algoritmo de árvores de decisão *Simple Cart* treinado em áreas previamente mapeadas localizadas no Estado do RS.

6.2 Material e métodos

Esse trabalho foi realizado em duas etapas. Para a realização da primeira etapa utilizou-se duas bacias hidrográficas localizadas em regiões distintas do Rio Grande do Sul e que apresentam diferenças no que se refere à área da bacia, ao relevo, ao material de origem, ao clima, e, conseqüentemente, aos solos, sendo elas a bacia do Santo Cristo e a do Arroio Portão. Nesta etapa foram realizados dois estudos independentes e semelhantes, um em cada bacia. As duas bacias foram sub-divididas em duas partes, sendo uma parte utilizada para o treinamento do modelo e a outra para a validação do modelo preditor. Para tal divisão buscou-se separar as bacias de tal modo que, em ambas as áreas de cada bacia ocorressem todas ou o máximo possível de unidades de mapeamento de solos (UMS), conforme apresentado na Figura 9.

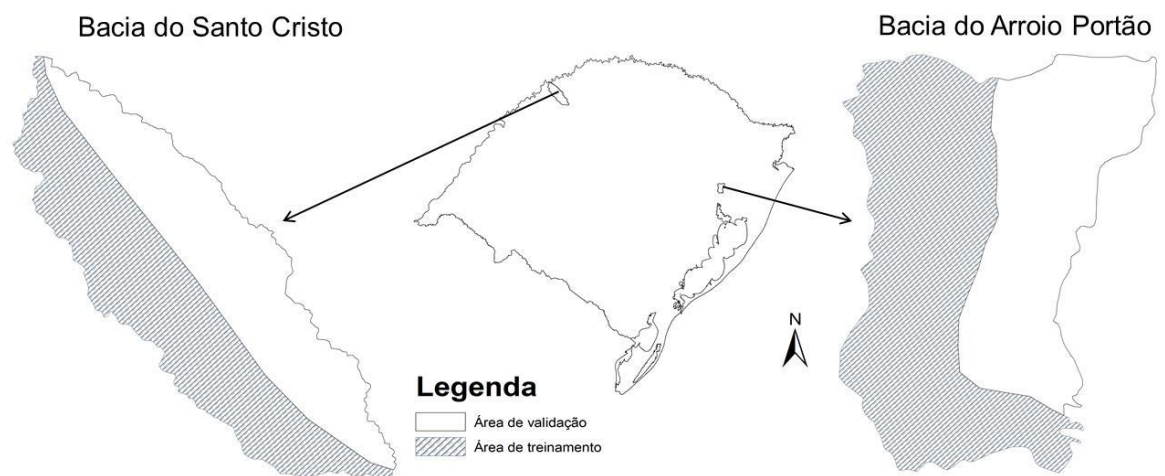


FIGURA 9. Figura esquemática da localização das áreas de treinamento e de validação das Bacias do Santo Cristo e do Arroio Portão.

Em ambiente SIG, utilizando o programa ArcGis 9.3 (ESRI, 2006), foram geradas as sete variáveis preditoras da distribuição dos solos na paisagem, sendo elas: elevação, declividade, comprimento de fluxo, acúmulo de fluxo, curvatura, índice de umidade topográfica e distância de rios. A amostragem de dados para treinamento dos algoritmos foi realizada de maneira aleatória com a densidade de amostragem de três pontos por hectare para ambas as bacias. Para a Bacia do Santo Cristo foram gerados 130.000 pontos amostrais e para a Bacia do Arroio Portão 30.000 pontos amostrais, nos quais foram amostrados valores de todas as variáveis preditoras citadas acima e da UM de solo ocorrente.

Para a realização da segunda etapa utilizou-se duas bacias localizadas numa mesma região, com características semelhantes de relevo, clima, material de origem e solos, sendo elas a bacia do Lageado Grande e a do Santo Cristo (item 3.1.1 e item 3.1.2). Para tanto, a bacia do Santo Cristo foi utilizada para a calibração do modelo e a do Lageado Grande para a validação do mesmo. Entretanto, como as duas bacias não apresentam a mesma legenda, fez-se necessário a adequação da legenda do mapa da bacia do Lageado Grande. Optou-se por adequar essa, pois a mesma possui um maior número de unidades de mapeamento de solos (15), enquanto, que a bacia do Santo Cristo possui 10. A partir disso elaborou-se um novo mapa com a finalidade de juntar as UMS reclassificadas (Figura 12). Esse ajuste foi realizado por um dos autores que participou no trabalho de campo e na elaboração dos mapas. A reclassificação foi realizada com base na similaridade das classes presentes na composição de cada UMS. Essa adequação está apresentada na Tabela 12.

A amostragem de dados para treinamento dos algoritmos foi realizada também de maneira simples e aleatória com a densidade de amostragem de três pontos por hectare na bacia do Santo Cristo, sendo o equivalente a 270.000 pontos amostrais, nos quais foram amostrados valores de todas as variáveis preditoras apresentadas no item 3.2 e da UM de solo ocorrente.

Após as coletas de dados, em ambas as etapas, os dados foram tabulados para a geração dos modelos e em seguida os mesmos foram aplicados no SIG para a geração dos mapas e posterior cálculos das acurácias conforme exposto no item 3.2.

TABELA 12. Legenda de identificação das Unidades de Mapeamento (UM) da bacia do Lageado Grande: simbologia usada no mapa de solos e a sua reclassificação de equivalência com a bacia do Santo Cristo. (OBS: As UM reclassificadas encontram-se sublinhadas).

UM	Composição da Unidade de Mapeamento	Reclassificação
	Lageado Grande	
LV1	Latossolo Vermelho	LV1
LV2	Associação de Latossolo Vermelho e Neossolo Regolítico	LV2
NV1	Nitossolo Vermelho	<u>LV1</u>
NV2	Associação de Nitossolo Vermelho e Neossolo Regolítico	<u>LV1</u>
M1	Chernossolo Argilúvico	M1
M2	Associação de Chernossolo e Neossolo Regolítico	<u>M1</u>
RR1	Neossolo Regolítico	RR1
RR2	Associação de Neossolo Regolítico e Neossolo Litólico, relevo forte ondulado	RR2
RR3	Associação de Neossolo Regolítico e Latossolo Vermelho	RR3
RR4	Associação de Neossolo Regolítico e Chernossolo Argilúvico	RR4
RR5	Associação de Neossolo Regolítico e Cambissolo e Nitossolo Vermelho	RR5
RR6	Associação Neossolo Regolítico e afloramento rochoso	<u>RR1</u>
RL	Associação de Neossolo Litólico e Neossolo Regolítico	RL
G	Complexo de Gleissolo Háplico/ Melânico	G

6.3 Resultados e discussão

Primeira etapa

Na Tabela 13 são apresentados os dados referentes a matriz de erro, a acurácia do usuário (AU), acurácia do mapeador (AM) e acurácia geral (AG) da área de treinamento e da área de validação para a bacia do Santo Cristo. Essa bacia é composta por 10 UMS, sendo que a UMS LV1 e LV2 estão associadas as partes da bacia que possuem relevo suave a ondulado e as demais estão associadas a relevo ondulado até fortemente ondulado.

Nesta tabela observa-se que o modelo preditivo foi capaz de reproduzir 63% do mapa da área de treinamento, entretanto quando realizada extrapolação dos dados a reprodutibilidade baixou para 50%. Observa-se também que, de modo geral, AM foi maior na área de treinamento, em

praticamente todas as UMS, quando comparadas à área de validação. Dentre as UMS preditas na área de treinamento, verifica-se que os maiores valores de AM estiveram associadas às UMS LV1 e RR1, com valores de AM de 74,3% e 71%, respectivamente. Isso quer dizer que as mesmas foram espacializadas de forma muito semelhante ao que se encontrará no mapa convencional de solos e que, ao consultar essas UMS no mapa, em 74,3% e 71% das vezes a indicação de ocorrência no mapa predito irá coincidir com a ocorrência no mapa original. A mesma tendência foi observada para a área de validação, porém com valores menores, de 58% para LV1 e 66,1% para RR1.

As UMS M1, RL, RR4 e RR5 não foram preditas pelos modelos. Isso, possivelmente ocorreu, devido à sua pequena extensão na área do mapeamento convencional de solos, de apenas 0,2%; 0,01%; 0,03% e 0,38%, respectivamente (Tabela 3). Contudo, percebe-se que essas UMS foram classificadas pelo modelo como sendo LV1 ou RR1. Isso ocorreu também com as demais UMS. Outro fator que pode ser observado em relação a essas UMS é que na área de treinamento elas aparecem com valores nulos, AM e AU como 0,0%, com exceção da M1.

Com relação às UMS RR4 e RR5, as mesmas não aparecem na área de treinamento porque as suas áreas estão localizadas somente na área de validação, portanto não foram inseridos dados para a construção do modelo. A RL aparece com informação nula pelo fato de que seus valores são baixos em relação aos demais, mas a mesma foi classificada como RR1 (0,001%) e RR2 (0,02%). Os mesmos resultados foram observados com a RR4 na área de validação, sendo que, ela também foi classificada como RR1 (0,04%) e como RR2 (0,01%). Na área de validação a RL não foi predita porque encontra-se somente na área de treinamento.

Assim, como os erros de classificação acabaram sendo inseridos nas UMS LV1 e RR1, ambas acabaram sendo superestimadas na área de treinamento, passando de 30% para 32,7% a LV1 e de 39,2% para 43,2% a RR1. Entretanto, na área de validação os erros de classificação foram inseridos, principalmente, na RR1 o que acarretou um aumento de aproximadamente 10% da sua área (30,1% para 40,5%) e, com isso, a área da LV1 diminuiu de 46,8% para 35,8%.

Os maiores valores de AM associados as classes LV1 e RR1 ocorreram devido a maior representatividade das mesmas no mapa convencional (38 e 35% da área total, respectivamente). Esse comportamento também foi observado em estudo realizado por Hengl et al. (2007) e Sarmiento (2010), nos quais, as classes predominantes na área dominaram no final das predições. Outro fator que pode ter favorecido, principalmente, quando se trata da UMS RR1 (Associação Neossolo e Cambissolos) está relacionado ao fato de que seis das sete variáveis utilizadas para a geração dos modelos são derivadas do relevo e o relevo tem forte influencia na formação, principalmente dos Neossolos, típicos de áreas com grande declividade.

Esses resultados também podem ser visualizados nas Figuras 10a e 10b, nas quais constata-se que a maior dificuldade de resposta do modelo ocorreu, principalmente, na parte centro-norte. Nessa região do mapa é onde ocorre a maior presença de UMS, bem como maior complexidade na distribuição das mesmas e, também, a região com relevo mais montanhoso e declivoso. Nesta figura também podemos observar que a UMS LV2 (representada pela cor vermelha), foi classificada, principalmente como RR1 (representada pela cor amarela). Contudo, esse erro de classificação pode estar associado ao fato de que, em ambas associações, ocorrem a presença de Neossolo Regolítico assim sendo, por essas UM se encontrarem em posições da paisagem muito semelhantes, a separação realizada no mapa convencional de solos não foi identificada pelo modelo. Outra UMS que também foi erroneamente predita pelo modelo, principalmente na área de validação foi a RR3 (representada pela cor azul). Nota-se que a mesma foi classificada no modelo no local onde no mapa convencional de solos a mesma não ocorre e, na área de ocorrência a mesma foi substituída pelas UMS LV1(verde), LV2 (vermelha), RR1(amarela) e RR2 (rosa). Entretanto, a RR3 é composta pela associação de Neossolo Regolítico e Latossolo Vermelho, as quais também se encontram na composição das UMS que a substituíram.

TABELA 13. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento e da área de validação da Bacia Santo Cristo

Área de Treinamento												
Mapa convencional												
Mapa predito	G	LV1	LV2	M1	RL	RR1	RR2	RR3	RR4	RR5	Total	AU (%)
G	0,2	0,2	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,5	41,1
LV1	0,4	22,3	2,6	0,0	0,0	6,1	0,7	0,8	0,0	0,0	32,7	68,0
LV2	0,1	1,0	3,1	0,0	0,0	0,7	0,2	0,1	0,0	0,0	5,2	59,3
M1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RL	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR1	1,0	5,8	3,0	0,1	0,0	27,9	3,8	1,7	0,0	0,0	43,2	64,5
RR2	0,1	0,6	1,5	0,0	0,0	4,1	7,4	1,3	0,0	0,0	15,0	49,5
RR3	0,0	0,2	0,1	0,0	0,0	0,6	0,5	2,0	0,0	0,0	3,4	59,5
RR4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Total	1,8	30,0	10,3	0,1	0,0	39,2	12,6	6,0	0,0	0,0	100,0	
AM (%)	11,8	74,3	29,9	0,0	0,0	71,0	58,7	33,9	0,0	0,0		
AG (%) = 63												
Área de Validação												
Mapa convencional												
Mapa predito	G	LV1	LV2	M1	RL	RR1	RR2	RR3	RR4	RR5	Total	AU (%)
G	0,5	0,5	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,4	36,4
LV1	0,5	27,2	0,8	0,0	0,0	4,2	0,0	2,9	0,0	0,3	35,8	76,0
LV2	0,1	3,3	0,3	0,0	0,0	0,9	0,2	1,6	0,0	0,0	6,5	4,8
M1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RL	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR1	1,8	10,7	2,4	0,1	0,0	19,9	1,6	3,3	0,0	0,5	40,5	49,2
RR2	0,3	4,6	0,9	0,1	0,0	3,9	1,5	1,8	0,0	0,0	13,1	11,4
RR3	0,0	0,6	0,5	0,0	0,0	1,2	0,1	0,3	0,0	0,0	2,8	9,6
RR4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Total	3,2	46,8	5,3	0,3	0,0	30,1	3,5	9,9	0,1	0,8	100,0	
AM (%)	15,5	58,0	5,9	0,0	0,0	66,1	42,9	2,7	0,0	0,0		
AG (%) = 50												

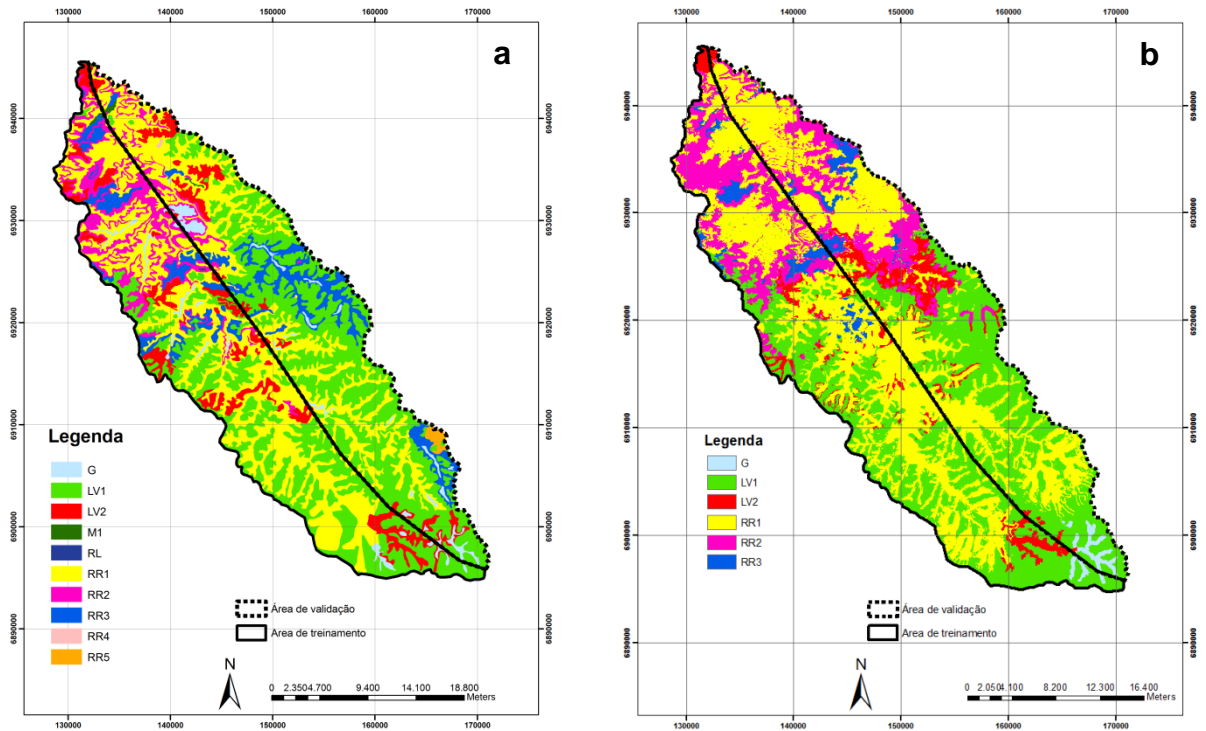


FIGURA 10. Mapa da bacia Santo Cristo com a distribuição original das classes de solos (a) e o mapa de solos predito (b).

Os resultados referentes a matriz de erro, AU, AM e AG da área de treinamento e da área de validação para a bacia do Arroio Portão são apresentados na Tabela 14.

Analisando os dados observa-se que, quanto à capacidade do modelo preditivo em reproduzir o mapa da área de treinamento, obteve-se uma reprodutibilidade de 68%, porém quando o modelo de AD extrapolou o mapeamento obteve-se a reprodutibilidade de 54%. Observa-se também que, de modo geral, AM foi maior na área de treinamento, em praticamente todas as UMS, quando comparadas à área de validação. Dentre as UMS preditas na área de treinamento verifica-se que, os maiores valores de AM estiveram associadas às UMS PVd, MXo/RLe e GMd, de 74,8%, 75% e 77,2%, respectivamente. Já em relação à área de validação as UMS que foram melhores preditas pelo modelo foram PVa/GMe e MXo/RLe com valores de 91,9% e 66,9%, respectivamente. As UMS CHa, SXd e aluvial, não foram preditas pelos modelos, possivelmente devido à sua pequena representatividade na área do mapeamento convencional de solos, de apenas 0,4; 0,6 e 0,3% da área, respectivamente. AS UMS CHa e aluvial aparecem

com valores nulos na área de treinamento porque estão situadas somente na área de validação.

Nesta bacia, as UMS MXo/RLe e CXf/RLe estão associadas as áreas de relevo mais declivoso, enquanto que as demais estão associadas a relevo plano a suave ondulado. Assim sendo, o modelo não conseguiu captar as diferenças para separar as UM associadas ao relevo menos declivoso, principalmente na separação entre as UM PVa e PVd que, na maioria dos casos, classificou como PVa/GMe. Isso pode ser observado, principalmente, na área de validação, na qual, a classe PVa, do seu total de 10,9% e 4,9% foram classificados corretamente pelo modelo, entretanto, 2,9% e 2,4% foram erroneamente classificados como PVd e como a associação PVa/GMe, respectivamente. Isso quer dizer que o modelo conseguiu prever o grande grupo Argissolo ou colocou em uma associação que continha a classe. O aumento dessa UM pode-se observar, principalmente na área de validação, onde ocorreu o aumento de área de 41,1% para 67,5%.

Esse fato, provavelmente, justifica-se pelo fato de que os solos predominantes em ambas UM pertencem a mesma classe de solo (Argissolo) e estão associadas ao mesmo tipo de relevo, portanto bastante semelhantes entre si em relação a variação dos atributos do terreno utilizados no trabalho. Resultados semelhantes foram observado por Figueiredo (2006), porém utilizando predição com regressões logísticas.

TABELA 14. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento e da área de validação da Bacia do Arroio Portão.

Área de treinamento											
Mapa convencional											
Mapa predito	CHa	PVd	PVa	MXo/RLe	SXd	Aluvial	CXf/RLe	GMd	PVa/GMe	Total	AU (%)
CHa	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
PVd	0,0	26,2	2,9	0,0	0,2	0,0	0,2	0,0	11,0	40,5	64,6
PVa	0,0	1,1	4,9	0,3	0,0	0,0	0,6	0,0	1,5	8,3	58,3
MXo/RLe	0,0	0,0	0,2	3,1	0,0	0,0	0,6	0,0	0,0	3,9	78,9
SXd	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Aluvial	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
CXf/RLe	0,0	0,1	0,5	0,7	0,0	0,0	1,7	0,0	0,4	3,4	49,1
GMd	0,0	0,0	0,0	0,0	0,0	0,0	0,0	3,9	0,7	4,6	84,8
PVa/GMe	0,0	7,5	2,4	0,0	0,1	0,0	0,2	1,1	27,9	39,2	71,1
Total	0,0	35,0	10,9	4,1	0,3	0,0	3,4	5,1	41,4	100,0	
AM (%)	0,0	74,8	44,7	75,0	0,0	0,0	50,2	77,2	67,4		
AG (%) = 68											
Área de validação											
Mapa convencional											
Mapa predito	CHa	PVd	PVa	MXo/RLe	SXd	Aluvial	CXf/RLe	GMd	PVa/GMe	Total	AU (%)
CHa	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
PVd	0,0	1,9	7,1	0,6	0,0	0,0	0,7	0,2	2,2	12,7	14,6
PVa	0,0	0,0	4,0	0,5	0,0	0,0	1,3	0,0	0,8	6,6	60,2
MXo/RLe	0,0	0,0	0,0	5,4	0,0	0,0	0,8	0,0	0,0	6,2	86,6
SXd	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Aluvial	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
CXf/RLe	0,0	0,0	0,3	1,6	0,0	0,0	2,1	0,0	0,0	4,0	52,9
GMd	0,0	0,2	0,0	0,0	0,0	0,0	0,0	2,5	0,2	2,9	84,9
PVa/GMe	1,1	5,8	11,9	0,0	1,2	0,8	0,7	8,3	37,8	67,5	55,9
Total	1,1	7,9	23,3	8,1	1,2	0,8	5,7	10,9	41,1	100,0	
AM (%)	0,0	23,6	17,0	66,9	0,0	0,0	37,6	22,5	91,9		
AG (%) = 54											

Esses resultados também podem ser visualizados nas figuras 11a e 11b, nas quais constata-se que a maior dificuldade de resposta do modelo ocorreu, principalmente, na região centro-sul. Percebe-se que o modelo não conseguiu separar a UMS PVa e classificando como PVa/GMe, UMS predominante, entretanto, a mesma está na composição da associação da qual foi substituída.

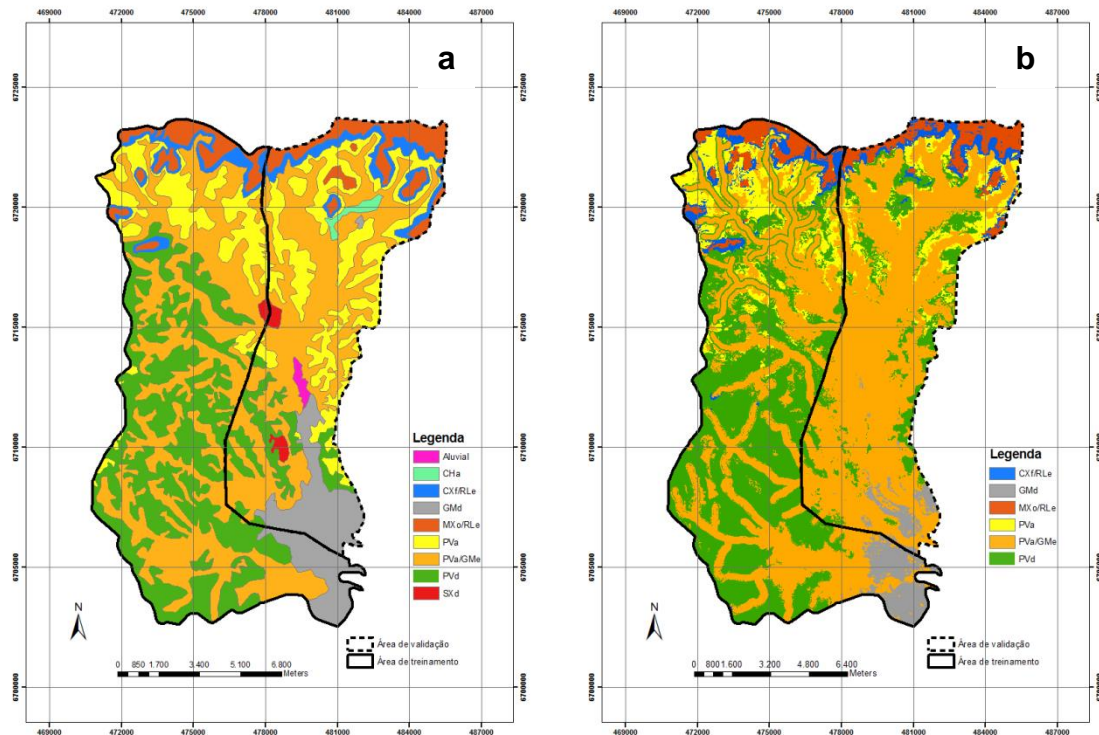


FIGURA 11. Mapa da bacia Arroio Portão com a distribuição original das classes de solos (a) e o mapa de solos predito (b).

Para as duas bacias observa-se que a AG foi maior na área de treinamento do que na área de validação, sendo a diferença entre elas de 13% menor na bacia do Santo Cristo e 14% na bacia do Arroio Portão. Grinand et al. (2008) observaram uma diferença da acurácia de 40% entre a área de treinamento e a área de validação quando da predição de classes de solo. ten Caten et. al. (2011a) utilizando regressão logística múltiplas, também observaram uma diferença de 22,5%, argumentando que a baixa acurácia na área de validação poderia ter origem na complexidade inerente à distribuição espacial do solo sobre a paisagem. Lemerrier et al. (2012), avaliando a predição da drenagem natural do solo e do material de origem, observaram diferenças de 18 e 24%, respectivamente entre área de treinamento e de validação. Entretanto, apesar das diferenças de AG entre a área de treinamento e a área de validação serem menores do que os valores encontrados por esses autores, os valores de AG encontrados nas áreas de extrapolação, 50% e 54%, foram superiores aos encontrados por Grinand et al. (2008) e ten Caten et. al. (2011), de 40% e 39%, respectivamente, contudo inferiores aos encontrados por Höfig et al. (2014), os quais obtiveram um valor

de AG de 75%, indicando assim que essa técnica ainda necessita de aprimoramento.

No presente trabalho, as diferenças da AG entre a área de treinamento e área de validação podem ter ocorrido devido a complexidade da distribuição do solo na paisagem, conforme observado por ten Caten et., (2011a), e também, devido às áreas não serem exatamente iguais, pois a distribuição das UMS bem como as feições do relevo, não se apresentam de forma igual nas duas metades da bacia. Assim sendo, considerando a complexidade da distribuição dos solos nas bacias, e que, de modo geral, em ambas as bacias, as acúcias nas UMS que apresentam maior extensão da área foram boas, pode-se utilizar esses resultados como uma ferramenta de auxílio ao mapeamento convencional.

Segunda etapa

Como os resultados da primeira etapa se mostraram promissores surgiu a necessidade de testar a técnica em duas bacias localizadas próximas, na mesma região e com características de relevo e clima semelhantes. Assim foi escolhido uma das bacias para a geração do modelo (bacia do Rio Santo Cristo) e a outra para a validação do modelo (bacia do Rio Lageado Grande). Os resultados dessa simulação são apresentados na tabela 15 e nas figuras 12a e 12b.

Observa-se nos resultados que o modelo foi capaz de predizer 61% do mapa da área de treinamento. A capacidade do modelo de reproduzir o mapa no qual o mesmo foi treinado foi semelhante ao encontrado nessa bacia no estudo realizado na primeira etapa (63%), entretanto, quando realizada a extrapolação para a validação na bacia do Lageado Grande a capacidade do modelo em reproduzir o mapa original baixou para 39%. ten Caten et al., (2011a) observaram resultados semelhantes em estudo realizado em uma bacia em São Pedro do Sul utilizando regressões logísticas múltiplas.

A avaliação da qualidade dos mapas gerados demonstra que para as UM LV1 e RR1 obteve-se o maior valor de AM, com um valor de aproximadamente 70%, tanto na área de treinamento como na área de validação. Esses valores mais elevados para essas UMS deve-se, basicamente, pelo fato de que as mesmas ocupam uma área mais

representativa do que as demais UM em ambas as bacias, com isso a quantidade de pontos usados na elaboração do modelo é maior que nas demais UMS e, assim, devido a baixa representatividade das demais UMS, o modelo não consegue captar toda a complexidade do ambiente de formação do solo das áreas a serem preditas pelos modelos (Hengl et al., 2007).

Entre as UMS preditas, verifica-se que a LV1 apresentou um valor de classificação correta de 70%, a RR1 de 60% e a RR2 de 10% do total. A LV2 apesar de ser predita pelo modelo, apresentou uma AM de 0%, ou seja, o modelo conseguiu predizer a mesma porém no local onde ela não ocorre no mapa original. Caso semelhante aconteceu com a RR3, apesar de uma pequena proporção ter sido classificada corretamente.

As demais UMS que não foram preditas foram incluídas, principalmente, nas UMS LV1 e RR1. Em função disso as UMS LV1 e RR1 foram superestimadas e a RR2 subestimada. A UMS LV1 passou de 29,4 % da área para 44,8%, a RR1 passou de 30,6% para 44,9%, e a RR2 diminuiu sua área de 24% para 4,1%, pois a mesma foi classificada como LV1 ou RR1.

A baixa capacidade de predição do modelo pode ter ocorrido pelo fato de que as UMS não serem compostas exatamente pelas mesmas classes de solos, são semelhantes, mas não iguais. Outro fator que pode ter determinado esses resultados é também a diferença de altitude, já que, a bacia do Lageado Grande está situada em uma altitude variando de 111 a 531 m, enquanto que a do Santo Cristo a altitude varia de 116 a 425 m. Estudo realizado Thomas et al., (1999) também constataram que a a altitude influenciou nos resultados obtidos, porém como as bacias são pequenas é difícil de avaliar a sua real influência. Juntamente com os fatores citados acima, independentemente das duas bacias estarem situadas numa região com o mesmo material de origem, os solos podem seguir uma mesma sequência de distribuição na paisagem, mas não em mesmas diferenças de altitude, o que também pode acabar influenciando nas diferenças de acurácia entre a área de calibração e a área de validação.

TABELA 15. Matriz de erros, acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG) da área de treinamento (bacia Santo Cristo) e da área de validação (bacia Lageado Grande).

Área de Treinamento – bacia Santo Cristo												
Mapa convencional												
Mapa predito	G	LV1	LV2	M1	RL	RR1	RR2	RR3	RR4	RR5	Total	AU (%)
G	0,3	0,1	0,1	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,5	0,5
LV1	0,6	28,4	2,1	0,0	0,0	6,6	1,0	3,0	0,0	0,3	41,9	0,7
LV2	0,2	0,7	2,0	0,0	0,0	0,6	0,3	0,1	0,0	0,0	3,9	0,5
M1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RL	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR1	1,1	7,9	3,1	0,1	0,0	25,0	3,2	2,3	0,0	0,1	42,9	0,6
RR2	0,1	0,4	0,3	0,0	0,0	1,9	3,5	0,5	0,0	0,0	6,9	0,5
RR3	0,2	0,5	0,3	0,0	0,0	0,7	0,3	2,0	0,0	0,0	4,0	0,5
RR4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Total	2,5	38,1	7,9	0,2	0,0	34,8	8,2	7,8	0,0	0,4	100,0	
AM (%)	0,1	0,7	0,3	0,0	0,0	0,7	0,4	0,3	0,0	0,0		
AG (%) = 61												
Área de Validação – bacia Lageado Grande												
Mapa convencional												
Mapa predito	G	LV1	LV2	M1	RL	RR1	RR2	RR3	RR4	RR5	Total	AU (%)
G	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
LV1	0,3	20,0	4,0	0,2	0,1	9,6	7,0	3,4	0,0	0,1	44,8	0,4
LV2	0,3	1,9	0,0	0,3	0,0	0,8	0,4	0,0	0,0	0,2	4,0	0,0
M1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RL	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR1	1,0	6,9	1,7	0,8	0,2	17,5	14,1	1,7	0,9	0,2	44,9	0,4
RR2	0,0	0,1	0,1	0,0	0,1	1,9	1,8	0,1	0,0	0,0	4,1	0,4
RR3	0,0	0,5	0,0	0,0	0,0	0,8	0,7	0,1	0,0	0,0	2,2	0,0
RR4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
RR5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Total	1,6	29,4	5,8	1,4	0,4	30,6	24,0	5,3	0,9	0,5	100,0	
AM (%)	0,0	0,7	0,0	0,0	0,0	0,6	0,1	0,0	0,0	0,0		
AG (%) = 39												

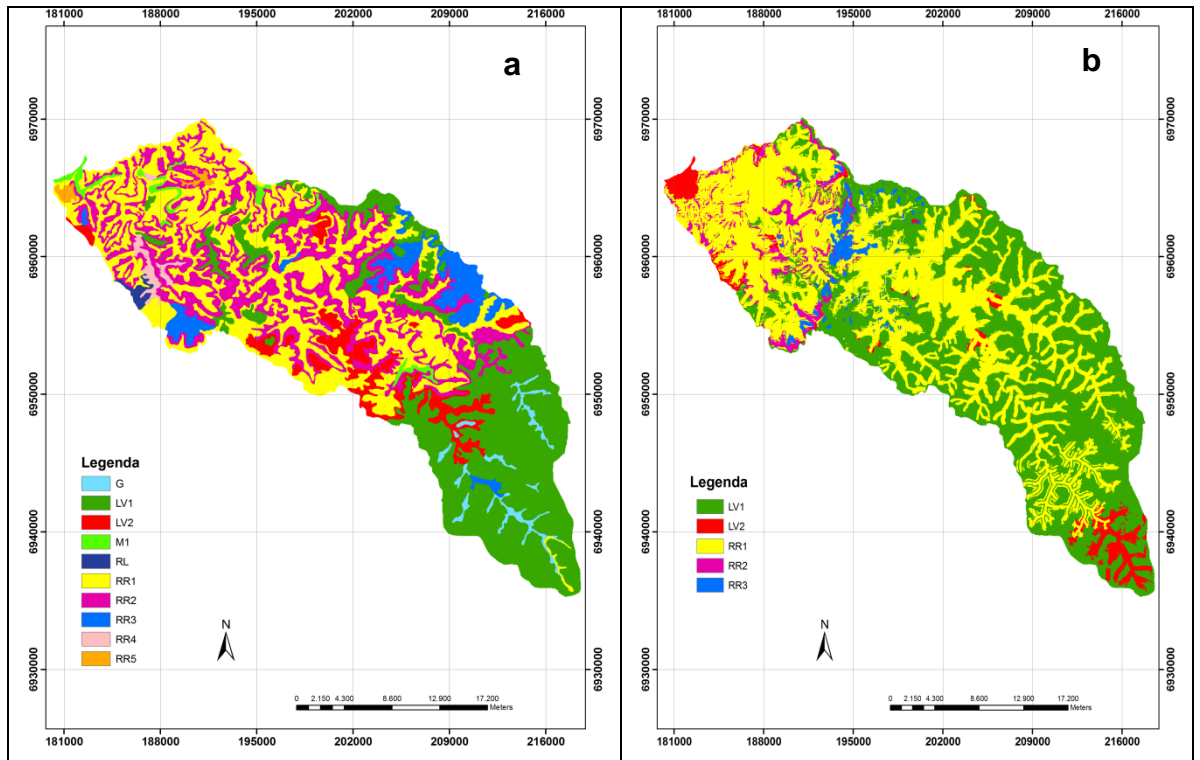


FIGURA 12. Mapa da bacia Lageado Grande com a distribuição original das classes de solos (a) e o mapa de solos predito (b).

6.4 Conclusões

A acurácia do mapeador e a acurácia geral encontradas foram maiores nas áreas de treinamento do que nas áreas de validação.

Os modelos superestimaram as unidades de mapeamento com maior representatividade nas bacias e subestimaram as de menor representatividade, tanto na área de treinamento como na de validação, em ambos estudos.

O mapeamento digital de solos a partir de uma área de treinamento pode ser utilizado como uma alternativa ou auxílio ao mapeamento convencional de solos.

Neste estudo, a utilização de árvores de decisão no mapeamento digital de solos se mostrou uma ferramenta promissora para elaboração de mapas vizinhos às áreas de referências e com mesma legenda, entretanto, não se mostrou tão eficiente para a elaboração de mapas para áreas mais distantes e com legendas parecidas.

7. CONCLUSÕES GERAIS

A presente tese avaliou e comparou o uso de diferentes métodos e materiais para o desenvolvimento e aplicação das técnicas do mapeamento digital de solos (MDS), com ênfase na predição de ocorrência de classes de solos.

No capítulo I avaliou-se o efeito da densidade de amostragem na acurácia dos modelos e verificou-se que quando não ocorreu a limitação do tamanho das árvores de decisão, o uso de densidades de amostragens maiores resultou em modelos com maior acurácia e com capacidade de prever um maior número de unidades de mapeamento de solos e que o uso de densidades de amostragem com menos de uma amostra por hectare se mostrou inadequado para a geração de mapas digitais de solos.

No capítulo II avaliou-se a influência de um conjunto de variáveis geomorfométricas na resposta dos modelos de predição. Neste estudo verificou-se que as variáveis comprimento de fluxo, elevação do terreno e distância dos rios exercem uma forte influência no resultado dos modelos de predição, tanto em relação à concordância com os mapas originais quanto ao número de unidades de mapeamento previstas. Os modelos gerados com somente essas variáveis geraram resultados de acurácia semelhantes aos modelos gerados com o conjunto de variáveis.

No capítulo III avaliou-se a capacidade de expansão por extrapolação do mapeamento de solos para áreas fisiograficamente semelhantes. Constatou-se nesse estudo que os modelos superestimaram as unidades de mapeamento com maior representatividade nas bacias e subestimaram as de menor representatividade. Constatou-se também que a utilização de árvores de decisão é uma ferramenta promissora para a expansão de mapas vizinhos e

com mesma legenda, entretanto não se mostrou tão eficientes na expansão para mapas próximos e com legendas parecidas.

Para trabalhos futuros sugere-se a aplicação dessas metodologias em áreas com características e/ou locais diferentes para testar e reprodução desses resultados, tanto ao que se refere à densidade de amostragem tanto ao conjunto de variáveis utilizadas e a expansão por extrapolação do mapeamento. Sugere-se também a utilização de modelos digitais de elevação com maiores resoluções para tentar separar classes de solos que se encontram numa mesma paisagem, mas em posições muito próximas na paisagem, como o caso dos Argissolos, na bacia do Arroio Portão e, além de testar-se as metodologias com mapas de solos convencionais com escalas maiores e menores. Para tentar melhorar a resposta dos modelos sugere-se a inserção de UMS com pouca representatividade em UMS de maior representatividade com características semelhantes.

8. REFERÊNCIAS

- ABRAMS, M.; HOOK, S.; RAMACHANDRAN, B. **ASTER users handbook**. Pasadena: JPL, 1999. 93p
- 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.
- BAILEY, N. et al. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. **Journal of Environmental Management**, London, v. 67, p. 395-407, 2003.
- BEHRENS, T. et al. Digital soil mappig using artificial neural networks. **Journal Plant Nutrition Soil Science**, Tenuco, v. 168, p. 1-13, 2005.
- BEHRENS, T.; SCHOLTEN, T. A comparison of data-mining techniques in predictive soil mapping. In: DIGITAL soil mapping, an introductory perspective. Developments in soil science. Amsterdam: Elsevier, 2007. p. 353-364.
- 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.
- BRASIL. Ministério da Agricultura. **Levantamento de reconhecimento dos solos do estado do Rio Grande do Sul**. Recife, 1973. 431p. (Boletim Técnico, 30).
- BREIMAN, L. et al. **Classification and regression trees**. Belmont, CA: Wadsworth International Group, 1984.
- BUI, E. N. Soil survey as a knowledge system. **Geoderma**, Amsterdam, v. 120, p. 17-26, 2004.
- 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 soi, survey over large spatial extents: na example from the Murray_Darling basin of Australia. **Geoderma**, Amsterdam, v. 111, p. 21-44, 2004.

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

CAVAZZI, S. et al. Are fine resolution digital elevation models always the best choice in digital soil mapping? **Geoderma**, Amsterdam, v. 195-196, p. 111-121, 2013.

CHAGAS, C. S. et al. Utilização de redes neurais artificiais para predição de classes de solo em uma bacia hidrografica no domínio de Mar dos Morros. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., Florianópolis, 2007. **Anais**. Florianópolis, 2007. p.2421-2428

CHAGAS, C. 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.

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.

COHEN, J. A coefficient of agreement for nominal scales. **Journal of Educational and Measurement**, Malden, v. 20, p. 37-46, 1960.

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.

CRIVELENTI, 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.

DOBOS, E. et al. Use of combined digital model and satellite radiometric data for regional soil mapping. **Geoderma**, Amsterdam, v. 97, p. 367-391, 2000.

ELNAGGAR, A. A. **Development of Predictive Mapping Techniques for Soil Survey and Salinity Mapping**. 2007. Thesis (PhD) - Oregon State University.

ESRI. **ArcGis 9.3**. [Software]. Redland, California, 2006.

FIGUEIREDO, S.R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. 2006. 92p. (Dissertação de Mestrado) – Programa de Pós-Graduação em

Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

FIGUEIREDO, S.R. et al. Uso de regressões logísticas múltiplas para mapeamento digital de solos no Planalto médio do RS. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, p. 2779-2785, 2008.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**, New York, v. 80, p. 185 – 201, 2002.

GAHEGAN, M. Is inductive machine learning just another wild goose (or might it lay the golden egg)? **International Journal of Geographic Information Science**, London, v. 17, n. 1, p. 69 – 92, 2003.

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 steeplands. **Scientia Agricola**, Piracicaba, v. 68, p. 167-174, 2011.

GIASSON, E. et al. Avaliação de cinco algoritmos de árvores de decisão e tres 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.

GERSHENFELD, N.A. **The Nature of Mathematical Modeling**. Cambridge University Press, 1999. 344 p.

GESSLER, P.E. et al. Modeling soil_landscape and ecosystem properties using terrain attributes. **Soil Science Society of American Journal**, Madison, v. 64, p. 2046-2056, 2000.

GRINAND, C. et al. Extrapolating regional soil landscape from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. **Geoderma**, Amsterdam, v. 143, p. 180-190, 2008.

HALL, M. et al. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v.11, 2009.

HASENACK, H.; WEBER, E. (Org.) **Base cartográfica vetorial contínua do Rio Grande do Sul - escala 1:50.000**. Porto Alegre: UFRGS-IB-Centro de Ecologia. 2010. 1 DVD-ROM (Série Geoprocessamento, 3)

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, 2008.

HENDERSON, B.L. et al. Australia-wide predictions of soil properties using decision trees. **Geoderma**, Amsterdam, v. 124, p. 383-398, 2005.

HENGL, T.; ROSSITER D.G. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. **Soil Science Society of America Journal**, Madison, v. 67, p. 1810-1822, 2003.

HENGL, T. et al. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, Amsterdam, v. 140, p. 417-427, 2007.

HJORT, J.; MARMION, M. **Effects of sample size on the accuracy of geomorphological models**. *Geomorphology*, Amsterdam, v. 102, p. 341 – 350, 2008.

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.

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.

KÄMPF, N.; GIASSON, E.; STRECK, E.V. **Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da microbacia do Rio Lageado Grande**. Porto Alegre: SEMA RS, 2004a. (Programa Nacional do Meio Ambiente II. Relatório final.)

KÄMPF, N.; GIASSON, E.; STRECK, E.V. **Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da microbacia do Rio Santo Cristo**. Porto Alegre: SEMA RS, 2004b. (Programa Nacional do Meio Ambiente II. Relatório final.)

KÄMPF, N.; CURI, N. **Formação e evolução do solo (pedogênese)**. In: PEDOLOGIA: fundamentos. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2012. p.207-302.

KEMPEN, B. et al. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. **Geoderma**, Amsterdam, v. 151, p. 311-326, 2009.

KEMPEN, B.; BRUS, D.J.; HEUVELINK, G.B.M. Soil type mapping using the generalised linear geo-statistical model: A case study in a Dutch cultivated peatland. **Geoderma**, Amsterdam, v.189-190, p. 540-553, 2012.

KER, J. C.; ALMEIDA, J. A. **Pedologia**. In: PROJETO RADAMBRASIL. Rio de Janeiro: IBGE, 1986. p. 405-791 (Levantamento de Recursos Naturais, v. 33).

KHEIR, R.B. et al. Spatial soil zinc content distribution from terrain parameters: A GIS-based decision-tree model in Lebanon. **Environmental Pollution**, Suíça, v.158, p.520-528, 2010.

KLAMT, E. et al. Solos da Bacia do Arroio Portão: Características, Distribuição e Aptidão de Uso. Porto Alegre: Departamento de Solos da UFRGS, 1992.

KIRKBY, M.J. **Hydrograph modeling strategies**. In: PROCESS in Physical and Human Geography. Oxford, UK, [S.n.],1975. p.69-90.

LAGACHERIE, P. et al. Mapping of reference area representativity using a mathematical soilscape distance. **Geoderma**, Amsterdam, v.101, p.105-118, 2001.

LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: DIGITAL Soil Mapping: an introductory perspective. Elsevier: Amsterdam, 2007. p. 3 - 24.

LARK, R.M.; BISHOP, T.F.A.; WEBSTER, R. Using expert knowledge with control of false discovery rate to select regressors for prediction of soil properties. **Geoderma**, Amsterdam, v. 138, p. 65–78, 2007.

LAWRENCE, R. et al. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. **Remote Sensing of Environment**, New York, v.90, p.331-336, 2004.

LEMERCIER, B. et al. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. **Geoderma**, Amsterdam, v.171-172, p.75-84, 2012.

LIPPITT, C. D. et al. Mapping Selective Logging in Mixed Deciduous Forest: A Comparison of Machine Learning Algorithms. **Photogrammetric Engineering and Remote Sensing**, v.74, p. 1201 – 1211, 2008.

MAcMILLAN, R.A.; MOON, D.E.; COUPÉ, R.A. Automated predictive ecological mapping in a Forest Region of B.C., Canada, 2001-2005. **Geoderma**, Amsterdam, v. 140, p. 353-373, 2007.

MCBRATNEY, A.B.; MENDONÇA SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, p. 3-52, 2003.

MILLER, J. FRANKLIN, J. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. **Ecological Modeling**, Amsterdam, v. 157, p. 227-247, 2001.

MINASNY, B.; McBRATNEY, A.B. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. **Geoderma**, Amsterdam, v. 142, p. 285-293, 2007.

MOORE, J.C. et al. Resource compartmentation and the stability of real ecosystems. **Nature**, London, v. 333, p. 261 - 263, 1988.

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.

NIELD, S.J.; BOETTINGER, J.L.; RAMSEY, R.D. Digitally mapping Gypsic and Natric soil areas using Landsat ETM data. **Soil Science Society of American Journal**, Madison, v.71, p.245-252, 2006.

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.

NOLASCO-CARVALHO, C. C.; 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, p. 46-55, 2009.

NELSON, M.A.; ODEH, O.A.; Digital soil class mapping using legacy soil profile data: a comparison of a genetic algorithm and classification tree approach. **Australian journal of soil research**, Clayton South, v.47, p.632-649, 2009.

ODGERS, N.P.; McBRATNEY, A.B.; MINASNY, B. Bottom-up digital soil mapping. I. Soil layer classes. **Geoderma**, Amsterdam, v. 163, p. 38-44, 2011.

OLAYA, V. Basic land-surface parameters. In: HENGL, T.; REUTER, H. (Ed.). **Geomorphometry, Concepts, Software, Application**. Elsevier, 2009, p. 141–169.

PAL, M.; MATHER, P. M. An assessment of the effectiveness of decision tree methods for land cover classification. **Remote Sensing of Environment**, New York, v. 86, p. 554 – 565, 2003.

PARK, S.J.; MCSWEENEY, K.; LOWERY, B. Identification of the spatial distribution of soils using a process-based terrain characterization. **Geoderma**, Amsterdam, v.103, p.249-272, 2001.

QI, F. Knowledge discovery from Area-Class resource maps: data preprocessing for noise reduction. **Transactions in GIS**, Malden, v.8, p.297-308, 2004.

QI, F.; ZHU, A-X. Knowledge discovery from soil maps using inductive learning. **International Journal of geographical information Science**, London, v.17, p.771-795, 2003.

ROMANO, N.; CHIRICO, G. B. **The role of terrain analysis in using and developing pedotransfer functions**. In: DEVELOPMENT of pedotransfer functions in soil hydrology. Amsterdam: [S.n.], 2004. p. 273 – 294.

ROSSITER, D.G. **Statistical methods for accuracy assessment of classified thematic maps**. Enschede: ITC, 2004. 46 p.

SARMENTO, E.C. **Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil**. 2010. 10p. (Dissertação de Mestrado) – Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

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.

SCHMIDT, K.; BEHRENS, T.; SCHOLTEN, T. Instance selection and classification tree analysis for large spatial datasets in digital soil mapping. **Geoderma**, Amsterdam, v. 146, p. 138-146, 2008.

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, M.P. et al. The effects of DEM resolution and neighborhood size on digital soil survey. **Geoderma**, Amsterdam, v. 137, p. 58-69, 2006.

SOROKINA, N.P.; KOZLOV, D.N. Experience in Digital Mapping of soil cover patterns. **Eurasian Soil science**, Moscou, v.42, p.198-210, 2009.

SUN, X.L. et al. Application of a digital soil mapping method in predicting soil orders on mountain areas of Hong Kong based on legacy soil data. **Pedosphere**, Nanjing, v. 21, p. 339–350, 2011.

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. ; PEDRON, F. A. ; Mendonça-Santos . Componentes principais como preditores no mapeamento digital de classes de solos. **Ciência Rural**, Santa Maria, v. 41, p. 1170-1176, 2011b.

TEN CATEN, A. et al. Regressões Logísticas múltiplas: Fatores que influenciam sua aplicação na predição de classes de solo, **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, p. 53-62, 2011c.

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. An appropriate data set size for digital soil mapping in Erechim, Rio Grande do Sul, Brazil. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 37, p. 359-366, 2013.

TESKE, R.; GIASSON, E.; BAGATINI, T. Comparação do uso de modelos digitais de elevação em mapeamento digital de solos em Dois Irmão, 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.

THOMAS, A.L. et al. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France). *Geoderma*, Amsterdam, v.90, p.291-305, 1999.

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

Yang, L. et al. Updating Conventional soil maps through Digital Soil Mapping, **Soil Science Society of America Journal**, Madison, v.75, p.1044-1053, 2011.

XU, M. et al. Decision tree regression for soft classification of remote sensing data. **Remote Sensing of Environment**, New York, v. 97, p. 322-336, 2005.

ZHU, A.X. et al. Automated soil inference under fuzzy logic. **Ecological Modelling**, Oxford, v. 90, p. 123–145, 1996.

ZHU, A.X. Mapping soil landscape as spatial continua: The neural network approach. **Water resources research**, Malden, v.36, p.663-677, 2000.

ZHU, A.X. et al. Soil mapping using GIS, expert knowledge, and fuzzy logic, **Soil Science Society American Journal**, London, v.65, p.1463-1472, 2001.

ZHOU, B.; ZHANG, X-G.; WANG, R-C. Automated soil resources mapping based on decision tree and bayesian predictive modeling. **Journal of Zhejiang University Science**, Hangzhou v.7, p.785-795, 2004.

9. APÊNDICES

Apêndice 1. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área interna da bacia do Lageado Grande, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 ponto por hectare					
2	20	39	0,45	59,2	5
10	22	43	0,45	59,2	5
50	18	35	0,41	57,9	4
100	9	17	0,41	57,8	3
200	4	7	0,41	57,5	3
300	4	7	0,41	57,5	3
0,3 ponto por hectare					
2	319	637	0,5	62,4	12
10	170	339	0,49	61,8	12
50	35	69	0,46	60,1	6
100	12	23	0,44	58,7	6
200	5	9	0,42	57,4	5
300	5	9	0,41	57,1	4
1 ponto por hectare					
2	1109	2217	0,58	68,5	14
10	696	1391	0,56	67	14
50	178	355	0,52	63,8	10
100	98	195	0,49	61,8	9
200	48	95	0,47	60,7	7
300	27	53	0,44	59,3	6
1,5 pontos por hectare					
2	2415	4829	0,6	69,7	14
10	1035	2069	0,58	68	14
50	253	505	0,53	64,5	14
100	138	275	0,51	63,2	10
200	70	139	0,48	61,3	9
300	35	69	0,45	60,2	6
2 pontos por hectare					
2	3192	6383	0,62	71,6	15
10	1161	2321	0,6	69,7	14
50	381	761	0,55	66	14
100	162	323	0,52	64,4	12
200	87	173	0,49	62,2	10
300	48	95	0,47	60,9	6
3 pontos por hectare					
2	8598	17195	0,67	74,8	15
10	2060	4119	0,62	71,7	15
50	495	989	0,57	67,3	14
100	248	495	0,54	65,5	14
200	162	323	0,52	63,9	13
300	95	189	0,49	62,4	12
4 pontos por hectare					
2	13151	26301	0,7	77,2	15
10	2836	5671	0,64	73,1	15
50	700	1399	0,58	68,4	14
100	375	749	0,55	66,5	14
200	210	419	0,53	64,8	12
300	127	253	0,51	63,3	12

Apêndice 2. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área maior da bacia do Lageado Grande, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 ponto por hectare					
2	20	39	0,45	59,2	5
10	22	43	0,45	59,2	5
50	18	35	0,41	57,8	4
100	9	17	0,41	57,8	3
200	4	7	0,4	57,5	3
300	4	7	0,41	57,7	3
0,3 ponto por hectare					
2	318	635	0,5	62,5	12
10	213	425	0,49	61,8	10
50	35	69	0,46	60,1	6
100	12	23	0,44	58,7	6
200	5	9	0,42	57,4	5
300	5	9	0,41	57,1	4
1 ponto por hectare					
2	1594	3187	0,58	68,3	14
10	601	1201	0,56	67,1	14
50	177	353	0,52	63,8	11
100	101	201	0,49	61,9	9
200	46	91	0,46	60,6	7
300	26	51	0,45	59,6	6
1,5 ponto por hectare					
2	2409	4817	0,6	69,7	14
10	1061	2121	0,57	67,9	14
50	257	513	0,53	64,5	11
100	140	279	0,51	63,2	10
200	74	147	0,48	61,3	7
300	36	71	0,45	60,2	6
2 ponto por hectare					
2	3252	6503	0,62	71,6	15
10	1149	2297	0,6	69,6	14
50	341	681	0,55	66,1	14
100	163	325	0,52	64,3	11
200	89	177	0,49	62,1	10
300	48	95	0,47	60,9	6
3 pontos por hectare					
2	8198	16395	0,67	74,8	15
10	2040	4079	0,62	71,7	14
50	518	1035	0,57	67,4	14
100	264	527	0,54	65,5	13
200	186	371	0,52	63,8	11
300	103	205	0,5	62,4	11
4 pontos por hectare					
2	13154	26307	0,7	77,2	15
10	2859	5717	0,64	73	15
50	678	1355	0,58	68,5	14
100	378	757	0,55	66,5	13
200	213	425	0,53	64,7	12
300	113	225	0,51	63,3	11

Apêndice 3. Resultados referentes ao número de folhas (NF), tamanho das árvores de decisão (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados normalizados na área maior da bacia do Lageado Grande, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 pontos por hectare					
2	73	145	0,46	60,3	7
10	34	67	0,46	60,3	6
50	16	31	0,45	59,2	5
100	7	13	0,44	58,7	4
200	4	7	0,44	58,6	4
300	4	7	0,41	57,9	4
0,3 pontos por hectare					
2	141	281	0,52	64,4	13
10	119	237	0,51	63,6	12
50	55	109	0,48	61,4	6
100	16	31	0,46	60,4	5
200	12	23	0,45	59,4	5
300	8	15	0,45	59	5
1 pontos por hectare					
2	1746	3491	0,6	69,6	15
10	544	1087	0,57	67,6	14
50	162	323	0,53	64,7	12
100	95	189	0,5	62,8	10
200	38	75	0,48	61,4	8
300	20	39	0,46	60,6	6
1,5 pontos por hectare					
2	4090	8179	0,62	71,3	15
10	1214	2427	0,59	69,3	15
50	254	507	0,54	65,8	14
100	133	265	0,52	64	11
200	67	133	0,49	62,2	9
300	34	67	0,47	61	8
2 pontos por hectare					
2	3366	6731	0,64	73,1	15
10	1291	2581	0,62	71,1	15
50	333	665	0,57	67,5	14
100	202	403	0,54	65,9	14
200	81	161	0,5	63,5	10
300	49	97	0,49	62,8	8
3 pontos por hectare					
2	9819	19637	0,68	76,1	15
10	2005	4009	0,64	72,8	15
50	463	925	0,58	68,5	14
100	271	541	0,56	66,7	13
200	142	283	0,53	64,6	12
300	99	197	0,51	63,5	11
4 pontos por hectare					
2	12677	25353	0,71	78,1	15
10	2739	5477	0,66	74	15
50	671	1341	0,59	69,4	14
100	318	635	0,57	67,5	13
200	183	365	0,54	65,5	12
300	129	257	0,53	64,6	12

Apêndice 4. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados normalizados na área maior da bacia do Santo Cristo, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 pontos por hectare					
2	104	207	0,43	61,4	6
10	85	169	0,43	61,4	6
50	31	61	0,4	59,3	5
100	20	30	0,36	58	5
200	17	33	0,33	56,1	5
300	2	3	0,25	52,4	4
0,3 pontos por hectare					
2	245	489	0,44	61,8	6
10	153	305	0,44	61,8	6
50	61	121	0,42	60,6	5
100	40	79	0,4	59,7	5
200	22	43	0,37	58,4	5
300	14	27	0,33	56,8	4
1 pontos por hectare					
2	1684	3367	0,52	66,8	8
10	982	1963	0,5	65,9	8
50	256	511	0,47	63,8	7
100	148	295	0,45	62,7	6
200	86	171	0,43	61,7	6
300	66	131	0,42	61,1	6
1,5 pontos por hectare					
2	7986	15971	0,55	68,8	10
10	1798	3595	0,52	67,1	9
50	464	927	0,48	64,4	8
100	235	469	0,45	63,0	7
200	119	237	0,44	62,1	6
300	82	163	0,43	61,6	6
2 pontos por hectare					
2	7332	14663	0,58	70,7	10
10	2071	4141	0,54	68,4	10
50	554	1107	0,49	65,3	8
100	294	587	0,47	63,8	8
200	143	285	0,44	62,7	6
300	108	215	0,44	62,0	6
3 pontos por hectare					
2	19970	39939	0,63	73,9	10
10	3485	6969	0,57	70,0	10
50	884	1767	0,51	66,2	8
100	448	895	0,48	64,6	8
200	226	451	0,46	63,4	7
300	137	273	0,45	62,7	6
400	112	223	0,44	62,4	6
4 pontos por hectare					
2	24788	49575	0,66	76,3	10
10	5975	11949	0,59	71,8	10
50	1399	2797	0,53	67,5	9
100	625	1249	0,50	65,8	8
200	303	605	0,47	64,2	8
300	178	355	0,46	63,4	6

Apêndice 5. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área interna da bacia do Santo Cristo, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 pontos por hectare					
2	150	299	0,41	60,3	6
10	103	205	0,41	60,2	6
50	28	55	0,38	58,6	5
100	14	27	0,35	57,3	4
200	11	21	0,3	55,1	3
300	3	5	0,25	52,1	2
0,3 pontos por hectare					
2	192	383	0,42	61,3	6
10	104	207	0,42	61,2	6
50	58	115	0,4	59,9	6
100	28	55	0,38	58,7	5
200	17	33	0,35	57,5	5
300	12	23	0,33	56,9	4
1 pontos por hectare					
2	2819	5637	0,52	67,1	10
10	994	1987	0,5	65,9	7
50	229	457	0,47	63,9	6
100	153	305	0,45	62,5	6
200	91	181	0,43	61,3	6
300	67	133	0,42	60,7	6
1,5 pontos por hectare					
2	4898	9795	0,55	68,7	10
10	1771	3541	0,52	67,0	10
50	456	911	0,47	64,3	8
100	238	475	0,45	63,0	6
200	98	195	0,44	61,6	6
300	86	171	0,43	61,0	6
2 pontos por hectare					
2	10487	20973	0,57	70,3	10
10	2401	4801	0,54	68,1	10
50	557	1113	0,49	64,8	8
100	291	581	0,46	63,4	6
200	133	265	0,44	62,3	6
300	111	221	0,44	61,7	6
3 pontos por hectare					
2	20666	41331	0,62	73,6	10
10	3687	7373	0,56	69,9	10
50	815	1629	0,5	66,1	9
100	430	859	0,48	64,5	8
200	201	401	0,46	63,3	7
300	132	263	0,45	62,5	6
4 pontos por hectare					
2	26121	52241	0,66	76,04	10
10	5711	11421	0,59	71,4	10
50	1205	2409	0,52	67,7	8
100	555	1109	0,5	65,6	8
200	318	635	0,47	64	8
300	193	385	0,46	63,2	6

Apêndice 6. Resultados referentes ao número de folhas (NF), tamanho das árvores (TA), kappa (K), acurácia geral (AG) e número de unidades de mapeamento preditas (UMS) referentes aos dados não normalizados na área maior da bacia do Santo Cristo, com diferentes números de elementos no nó final (M)

M	NF	TA	K	AG, %	UMS
0,1 pontos por hectare					
2	150	299	0,41	60,3	6
10	103	205	0,41	60,2	6
50	28	55	0,38	58,6	5
100	14	27	0,35	57,3	4
200	11	21	0,3	55,1	3
300	3	5	0,25	52,1	2
0,3 pontos por hectare					
2	192	383	0,42	61,3	6
10	104	207	0,42	61,2	6
50	58	115	0,4	59,9	6
100	28	55	0,38	58,7	5
200	17	33	0,35	57,5	5
300	12	23	0,33	56,9	4
1 pontos por hectare					
2	2819	5637	0,52	67,1	10
10	994	1987	0,5	65,9	7
50	229	457	0,47	63,9	6
100	153	305	0,45	62,5	6
200	91	181	0,43	61,3	6
300	67	133	0,42	60,7	6
1,5 pontos por hectare					
2	4898	9795	0,55	68,7	10
10	1771	3541	0,52	67,0	10
50	456	911	0,47	64,3	8
100	238	475	0,45	63,0	6
200	98	195	0,44	61,6	6
300	86	171	0,43	61,0	6
2 pontos por hectare					
2	10487	20973	0,57	70,3	10
10	2401	4801	0,54	68,1	10
50	557	1113	0,49	64,8	8
100	291	581	0,46	63,4	6
200	133	265	0,44	62,3	6
300	111	221	0,44	61,7	6
3 pontos por hectare					
2	20666	41331	0,62	73,6	10
10	3687	7373	0,56	69,9	10
50	815	1629	0,5	66,1	9
100	430	859	0,48	64,5	8
200	201	401	0,46	63,3	7
300	132	263	0,45	62,5	6
4 pontos por hectare					
2	26121	52241	0,66	76,04	10
10	5711	11421	0,59	71,4	10
50	1205	2409	0,52	67,7	8
100	555	1109	0,5	65,6	8
200	318	635	0,47	64	8
300	193	385	0,46	63,2	6