

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

COMPARAÇÃO ENTRE QUATRO ALGORITMOS DE APRENDIZAGEM DE
MÁQUINA NO MAPEAMENTO DIGITAL DE SOLOS NO VALE DOS
VINHEDOS, RS, BRASIL

Eliana Casco Sarmento
(Dissertação de Mestrado)

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

**COMPARAÇÃO ENTRE QUATRO ALGORITMOS DE APRENDIZAGEM DE
MÁQUINA NO MAPEAMENTO DIGITAL DE SOLOS NO VALE DOS
VINHEDOS, RS, BRASIL**

Eliana Casco Sarmiento
Engenheira Agrônoma e Bióloga (UFRGS)

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

CIP - CATALOGAÇÃO INTERNACIONAL NA PUBLICAÇÃO
Biblioteca Setorial da Faculdade de Agronomia da UFRGS

S246c Sarmento, Eliana Casco
Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil / Eliana Casco Sarmento — Porto Alegre : E.C. Sarmento, 2010.

xv, 109f; II.

Dissertação(Mestrado) – Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

1. Solo : Vale dos Vinhedos : Rio Grande do Sul : Mapeamento Digital. I. Título.

CDD: 631.4

ELIANA CASCO SARMENTO
Engenheira Agrônoma - UFRGS
Bacharel em Ciências Biológicas - UFRGS

DISSERTAÇÃO

Submetida como parte dos requisitos
para obtenção do Grau de

MESTRE 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), Brasil

Aprovado em: 31.03.2010
Pela Banca Examinadora

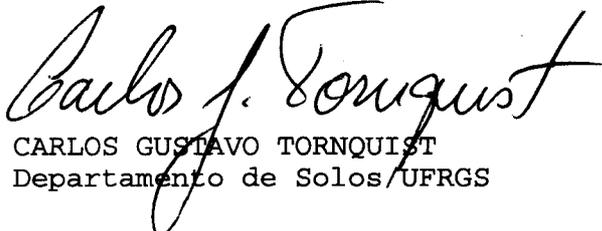
Homologado em: 09.08.2010
Por



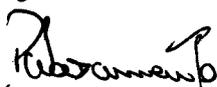
ELVIO GIASSON
Orientador-PPG Ciência do Solo



FLAVIO A. O. CAMARGO
Coordenador do
Programa de Pós-Graduação em
Ciência do Solo



CARLOS GUSTAVO TORNQUIST
Departamento de Solos/UFRGS



PAULO CÉSAR DO NASCIMENTO
Departamento de Solos/UFRGS



ULISSES FRANZ BREMER
Departamento de Geografia/UFRGS



PEDRO ALBERTO SELBACH
Diretor da Faculdade
de Agronomia

Dedico a Eliseu, meu companheiro epicurista.

AGRADECIMENTOS

Aos meus pais, Nelson (*in memoriam*) e Paulina, pelo carinho, apoio, ensinamentos e a minha própria existência;

Ao meu orientador, prof. Élvio Giasson, pela orientação e contribuição em temas essenciais, bem como pela confiança depositada e liberdade proporcionada na escolha do tema de pesquisa;

Ao PPG em Ciência do Solo, bem como aos professores e funcionários pela oportunidade de desenvolvimento pessoal e profissional;

Ao CNPq, pela concessão da bolsa de estudos que permitiu total dedicação ao curso e à elaboração desta dissertação;

Ao pesquisador Carlos Alberto Flores, da Embrapa Clima Temperado, pela gentil cedência do mapa detalhado de solos da área de estudo e constante troca de experiências;

Ao laboratório de geoprocessamento do Centro de Ecologia da UFRGS, na pessoa do prof. Heinrich Hasenack, pela amizade, estímulo e oportunidade de participação em diversos projetos, entre eles o projeto APL Vinhos;

Ao pesquisador Jorge Tonietto, da Embrapa Uva e Vinho, pela oportunidade de participação no projeto APL Vinhos;

À FINEP, pelo financiamento do projeto APL Vinhos, cujos dados cartográficos e de solos serviram de base para esta pesquisa.

COMPARAÇÃO ENTRE QUATRO ALGORITMOS DE APRENDIZAGEM DE MÁQUINA NO MAPEAMENTO DIGITAL DE SOLOS NO VALE DOS VINHEDOS, RS, BRASIL^{1/}

Autor: Eliana Casco Sarmiento
Orientador: Prof. Élvio Giasson

RESUMO

A carência de dados de solos estimulou o desenvolvimento de técnicas de modelagem digital para estimar classes ou propriedades dos solos. Este trabalho compara quatro algoritmos de aprendizagem de máquina, três redes neurais (Fuzzy ARTMap, SOM e MLP) e uma árvore de decisão (Gini), na predição de ordens de solos no Vale dos Vinhedos, Rio Grande do Sul, Brasil. O material utilizado inclui um Modelo Numérico do Terreno (MNT) com resolução de 5 metros, uma base cartográfica digital em escala 1:5.000, um mapa convencional detalhado de solos e o *software* de Sistema de Informação Geográfica (SIG) Idrisi. A partir do MNT e da base foram calculadas sete variáveis topográficas e hidrológicas, cujos valores e a ordem de solo foram lidos em pontos aleatoriamente distribuídos com cinco densidades de amostragem: 0,5, 1, 1,5, 2 e 4 pontos.ha⁻¹. Os dados obtidos nos pontos foram empregados para treinar os algoritmos e classificar as ordens de solos no restante da área de estudo. A avaliação de acurácia foi realizada através de matriz de erros, exatidão geral e índice *Kappa*, tomando o mapa convencional como referência. A árvore de decisão obteve melhor exatidão geral (71%) e índice *Kappa* (0,58), com 2 pontos.ha⁻¹, e apresentou menor sensibilidade à densidade de amostragem, obtendo índice *Kappa* acima de 0,5 a partir de 1 ponto.ha⁻¹. A rede neural MLP alcançou desempenho próximo à árvore de decisão, mas foi mais sensível à densidade de amostragem, obtendo índice *Kappa* acima de 0,5 a partir de 1,5 pontos.ha⁻¹. A árvore de decisão Gini e a rede neural MLP mostraram potencial para a predição de ordens de solos em escala de detalhe.

1/ Dissertação de Mestrado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (109 p.) Março, 2010. Trabalho realizado com apoio financeiro da FINEP e do CNPq.

COMPARISON OF FOUR MACHINE LEARNING ALGORITHMS FOR DIGITAL SOIL MAPPING IN THE VALE DOS VINHEDOS, RS, BRAZIL^{2/}

Author: Eliana Casco Sarmiento

Adviser: Prof. Élvio Giasson

ABSTRACT

The lack of spatial soil data has stimulated the development of modeling techniques for predicting soil classes or soils properties. This work compares four machine learning algorithms, three neural networks (Fuzzy ARTMap, SOM, and MLP) and a decision tree (Gini), to predict soil classes in the State Rio Grande do Sul, Brazil. A digital elevation model (DEM) with spatial resolution of 5 meters, a digital cartographic base at 1:5,000 scale, and a detailed conventional soil map were used to run the models on Idrisi GIS software. Seven topographic and hydrological variables were derived from the DEM and the cartographic base, and values were read in randomly distributed sampling points with five sampling densities: 0.5, 1, 1.5, 2 and 4 sampling points per hectare. Data read at sample points were used to train the algorithms and to estimate soil classes in the whole study area. Accuracy was evaluated using error matrices, overall accuracy and *Kappa* Index, using the conventional soil map as reference for comparison. The best overall accuracy (71%) and *Kappa* Index (0.58) were obtained by the Gini decision tree with sampling density of 2 points per hectare. Decision tree was less sensitive to sampling density, obtaining *Kappa* Index above 0.5 in all sampling densities equal or higher than 1 point per hectare. Neural network MLP (*Multi-layer Perceptron*) showed a similar performance, but was more sensitive to sampling density, obtaining *Kappa* Index above 0.5 only with sampling densities equal or higher than 1.5 points per hectare. Gini decision tree and neural network MLP showed potential to predict soil class at detailed scale.

2/ M.Sc. Dissertation in Soil Science. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (109 p.) March, 2010.

SUMÁRIO

	Página
1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	3
2.1. Levantamentos de solos.....	3
2.2. Mapeamento convencional de solos	4
2.3. Mapeamento digital de solos.....	7
2.3.1. Principais métodos e técnicas	11
2.3.1.1. Regressões múltiplas.....	11
2.3.1.2. Geoestatística	13
2.3.1.3. Lógica de conjuntos difusos (lógica <i>fuzzy</i>).....	17
2.3.1.4. Redes neurais artificiais	19
2.3.1.5. Árvores de decisão	22
2.3.2. Uso de atributos do terreno	26
2.3.2.1. Declividade	27
2.3.2.2. Orientação	28
2.3.2.3. Curvatura	28
2.3.2.4. Direção de fluxo	28
2.3.2.5. Acúmulo de fluxo	29
2.3.3. Avaliação da acurácia de mapas	29
2.3.3.1. Matriz de erro.....	30
2.3.3.2. Exatidão geral	32
2.3.3.3. Índice <i>Kappa</i>	33
2.3.4. Questões relacionadas à avaliação de acurácia.....	34
3. MATERIAL E MÉTODOS.....	36
3.1. Caracterização da área de estudo.....	36
3.1.1. Localização.....	36
3.1.2. Clima.....	38
3.1.3. Geologia	40
3.1.4. Relevo.....	41
3.1.5. Solos.....	43
3.1.6. Vegetação.....	45
3.2. Material utilizado.....	46
3.2.1. Mapa convencional de solos.....	46

3.2.2. Base cartográfica digital.....	51
3.2.3. Modelo Numérico do Terreno	52
3.2.4. <i>Softwares</i>	54
3.3. Metodologia.....	54
3.3.1. Agrupamento do mapa de solos	55
3.3.2. Geração dos pontos de amostragem.....	57
3.3.3. Cálculo das variáveis preditoras	60
3.3.3.1. Elevação (MNT)	60
3.3.3.2. Declividade	61
3.3.3.3. Orientação	63
3.3.3.4. Curvatura	64
3.3.3.5. Direção de fluxo	65
3.3.3.6. Acúmulo de fluxo	66
3.3.3.7. Distância da rede de drenagem	67
3.3.4. Treinamento dos algoritmos de classificação e predição da ocorrência das ordens de solos.....	68
3.3.4.1. Rede neural <i>Multi-layer Perceptron</i> (MLP).....	69
3.3.4.2. Rede neural <i>Adaptive Resonance Theory</i> (ART).....	70
3.3.4.3. Rede neural <i>Self-Organizing Map</i> (SOM)	71
3.3.4.4. Árvore de decisão	73
3.3.5. Avaliação da acurácia.....	74
4. RESULTADOS E DISCUSSÃO	75
4.1. Mapas estimados	75
4.2. Matrizes de erro.....	77
4.3. Exatidão geral.....	86
4.4. Índice <i>Kappa</i>	87
4.5. Discussão geral	90
4.5.1. Sobre as variáveis preditoras	90
4.5.2. Sobre os mapas estimados e o mapa convencional de solos	92
4.5.3. Sobre as densidades de amostragem	93
4.5.4. Sobre os algoritmos de classificação.....	94
5. CONCLUSÕES	97
6. REFERÊNCIAS BIBLIOGRÁFICAS	99

RELAÇÃO DE TABELAS

	Página
Tabela 1. Esquema de uma matriz de erros para caso com três categorias (A, B e C). \sum_e = Total estimado, EI = erros de inclusão, AU = acurácia do usuário, \sum_r = total da referência, EO = erros de omissão, AM = acurácia do mapeador.....	31
Tabela 2. Superfície ocupada por diferentes fases de relevo na área de estudo.	42
Tabela 3. Sigla, descrição da unidade de mapeamento, número de manchas mapeadas, área em hectares e proporção ocupada por unidade de mapeamento de solos na área de estudo (Flores, 2006c).	48
Tabela 4. Área em hectares e proporção ocupada na área de estudo pelos solos agrupados para o primeiro nível (Ordem).	57
Tabela 5. Densidade de amostragem, número total de pontos e número de pontos por ordem de solos.	58
Tabela 6. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 0,5 ponto.ha ⁻¹	78
Tabela 7. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 0,5 ponto.ha ⁻¹	78
Tabela 8. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 0,5 ponto.ha ⁻¹	78
Tabela 9. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 0,5 ponto.ha ⁻¹	79
Tabela 10. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 1 ponto.ha ⁻¹	79
Tabela 11. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 1 ponto.ha ⁻¹	79
Tabela 12. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 1 ponto.ha ⁻¹	80
Tabela 13. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 1 ponto.ha ⁻¹	80
Tabela 14. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 1,5 pontos.ha ⁻¹	80
Tabela 15. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 1,5 pontos.ha ⁻¹	81
Tabela 16. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 1,5 pontos.ha ⁻¹	81
Tabela 17. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 1,5 pontos.ha ⁻¹	81

Tabela 18. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 2 pontos.ha ⁻¹	82
Tabela 19. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 2 pontos.ha ⁻¹	82
Tabela 20. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 2 pontos.ha ⁻¹	82
Tabela 21. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 2 pontos.ha ⁻¹	83
Tabela 22. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 4 pontos.ha ⁻¹	83
Tabela 23. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 4 pontos.ha ⁻¹	83
Tabela 24. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 4 pontos.ha ⁻¹	84
Tabela 25. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 4 pontos.ha ⁻¹	84
Tabela 26. Exatidão geral (%) dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.....	86
Tabela 27. Índice <i>Kappa</i> dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.....	88
Tabela 28. Matriz de erro entre o mapa estimado com a árvore de decisão Gini com densidade de amostragem de 2 pontos.ha ⁻¹ (colunas) e o mapa estimado com a rede neural MLP com densidade de amostragem de 4 pontos.ha ⁻¹ (linhas).	90

RELAÇÃO DE FIGURAS

	Página
Figura 1. Exemplo de um variograma (Hengl, 2007).	15
Figura 2. Esquema de mapeamento digital de solos com <i>regressão-kriging</i> (Dobos et al., 2006).....	16
Figura 3. Exemplo de função de pertinência a conjunto difuso em contraste com uma função de conjunto booleano.	17
Figura 4. Esquema de uma rede neural com uma camada oculta (diferentes cores das linhas indicam conexões com pesos diferentes entre nós).....	20
Figura 5. Esquema de uma árvore de decisão com quatro variáveis e três classes. Xi representa as variáveis; a, b, c, d, e representam os limites da regra de decisão; A, B, C representam as classes (Zhou et al., 2004).....	23
Figura 6. (a) Esquema da determinação da direção do fluxo sobre um MNT (Li et al., 2005); (b) Representação da direção de fluxo sobre uma área maior (Pennock, 2003).....	29
Figura 7. Localização da região de produção vitivinícola da Serra Gaúcha e do Vale dos Vinhedos no Rio Grande do Sul.	37
Figura 8. Localização da área de estudo.	38
Figura 9. Temperatura média mensal e precipitação pluvial total mensal em Bento Gonçalves, RS (1961 – 1990).	39
Figura 10. Vista geral do relevo da região.	42
Figura 11. Vista em perspectiva tridimensional da área de estudo, de noroeste para sudeste.	43
Figura 12. Padrão de uso do solo da região, com exploração agrícola nas áreas menos íngremes (várzeas e patamares dos derrames) e remanescentes de florestas nativas nas áreas mais acidentadas.	46
Figura 13. Mapa convencional de solos disponível para a área de estudo (Sarmiento et al., 2008).	47
Figura 14. Rede de drenagem superficial e curvas de nível com equidistância vertical de 5 metros, extraídas da base cartográfica digital disponível para a área de estudo (Sarmiento et al., 2008).....	52
Figura 15. Visualização em perspectiva do MNT da área de estudo, vista de sudoeste para nordeste.....	53
Figura 16. Fluxograma com os principais passos da metodologia utilizada.....	55

Figura 17. Mapa convencional de solos, com 36 unidades de mapeamento, e mapa agrupado para o primeiro nível (Ordem), com 4 classes.	57
Figura 18. Mapa de solos simplificado sobreposto com os pontos das cinco densidades de amostragem.	59
Figura 19. Mapa de elevação (MNT) da área de estudo.	61
Figura 20. Mapa de declividade da área de estudo, expressa em percentagem.	62
Figura 21. Mapa de orientação das encostas na área de estudo.	63
Figura 22. Mapa de curvatura da área de estudo.	64
Figura 23. Mapa de direção de fluxo da área de estudo.	65
Figura 24. Mapa de acúmulo de fluxo da área de estudo.	67
Figura 25. Mapa de distâncias da rede de drenagem na área de estudo.	68
Figura 26. Parâmetros de treinamento da rede neural MLP com densidade de 1 ponto por hectare.	70
Figura 27. Parâmetros de treinamento da rede neural Fuzzy ARTMap com densidade de 1 ponto por hectare.	71
Figura 28. Parâmetros de treinamento da rede neural SOM com densidade de 1 ponto por hectare.	72
Figura 29. Parâmetros de treinamento da árvore de decisão Gini com densidade de 1 ponto por hectare.	74
Figura 30. Mapas estimados através dos quatro algoritmos de aprendizagem de máquina e cinco densidades de amostragem: (a) 0,5 pontos.ha ⁻¹ ; (b) 1 ponto.ha ⁻¹ ; (c) 1,5 pontos.ha ⁻¹ ; (d) 2 pontos.ha ⁻¹ ; (e) 4 pontos.ha ⁻¹	76
Figura 31. Exatidão geral dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.	87
Figura 32. Índice <i>Kappa</i> dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.	88

RELAÇÃO DE ABREVIATURAS

APL	Arranjos Produtivos Locais
ART	<i>Adaptive Resonance Theory</i>
CAD	<i>Computer Aided Design</i>
CENECO	Centro de Ecologia
DO	Denominação de Origem
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAO	<i>Food and Agriculture Organization</i>
FINEP	Financiadora de Estudos e Projetos
GPS	<i>Global Positioning System</i>
MAXVER	Máxima Verossimilhança
MLA	<i>Machine Learning Algorithms</i>
MLP	<i>Multi-Layer Perceptron</i>
MNT	Modelo Numérico do Terreno
SIG	Sistema de Informação Geográfica
SOM	<i>Self-Organizing Map</i>
TIN	<i>Triangulated Irregular Network</i>
UCS	Universidade de Caxias do Sul
UFRGS	Universidade Federal do Rio Grande do Sul
UTM	Universal Transverso de Mercator

1. INTRODUÇÃO

Os levantamentos convencionais de solos contemplam um estudo do terreno e das principais características dos perfis de solo, compreendendo a descrição morfológica, a caracterização física e química e a classificação taxonômica, além da espacialização dos limites das unidades cartográficas para a obtenção do mapa final de classes de solo. Pela intensidade de trabalho de campo e de análises em laboratório, tais levantamentos resultam demorados e dispendiosos, o que limita sua elaboração e explica a escassez quase generalizada de mapas de solos.

De outro lado, a pressão por recursos exercida pelas necessidades crescentes da população humana e a emergência de diversos temas relacionados a essa questão tem incrementado a demanda por informações de solos. Aplicações como planejamento da produção agrícola, elaboração de zoneamentos, avaliações ambientais, conservação de recursos hídricos, expansão urbana, disposição de resíduos e estudos sobre mudanças climáticas, entre outros, necessitam de mapas de solos como subsídio para tomadas de decisão.

A indisponibilidade de mapas com qualidade ou detalhamento suficiente fomentou o desenvolvimento de novas técnicas de modelagem para prever e espacializar propriedades ou tipos de solos. A difusão de tecnologias como o Sistema de Posicionamento Global (GPS), imagens de sensoriamento remoto e os Sistemas de Informação Geográfica (SIG), facilitou a integração de informações espaciais de diferentes origens e a realização de análises complexas, possibilitando a emergência do mapeamento digital de solos (Morris, et al., 2000; Zhu et al., 2001; McBratney et al., 2003; Giasson et al., 2006; Brodsky et al., 2006). Essa nova área da Ciência do Solo assume importância pelo seu potencial em produzir informações digitais de solos de forma relativamente rápida e com medidas quantitativas de qualidade.

Assim como nos levantamentos convencionais, grande parte dos estudos em mapeamento digital de solos baseia-se nas relações entre os solos e as características da paisagem. Entretanto, enquanto nos primeiros as relações solo-paisagem são estabelecidas de forma qualitativa, no mapeamento digital elas são avaliadas quantitativamente a partir de variáveis

derivadas de Modelos Numéricos do Terreno (MNT). As abordagens mais empregadas para quantificar essas relações se baseiam em métodos paramétricos, entre os quais regressões lineares, regressões logísticas, análises geoestatísticas e lógica de conjuntos difusos (lógica *Fuzzy*). Sua aplicação exige uma base de conhecimento consistente sobre solos e sobre os processos pedogenéticos e suas relações com a paisagem.

Mais recentemente métodos não paramétricos baseados em algoritmos de aprendizagem de máquina (*Machine learning algorithms* - MLA) também têm sido aplicados ao mapeamento digital de solos (Grinand et al., 2008). Aprendizagem de máquina é uma abordagem computacional de particionamento e categorização de dados que se baseia na idéia de “aprender” padrões existentes em bases de dados. Entre as vantagens cita-se a capacidade de processar grandes volumes de dados multidimensionais com baixo nível de intervenção humana, tempo de processamento reduzido e acurácia equivalente ou superior a métodos paramétricos (Lippitt et al., 2008). Todavia, ainda há uma compreensão limitada desses algoritmos em função de uma falta de conhecimento sobre suas capacidades, limitações e potencial de aplicação para o mapeamento digital de solos.

Este trabalho compara quatro algoritmos de aprendizagem de máquina para o mapeamento digital de solos, utilizando um mapa detalhado de solos como referência. O estudo baseia-se na hipótese de que é possível estimar ordens de solos a partir de variáveis descritoras do relevo em regiões onde este é um fator de formação preponderante. O objetivo principal é avaliar o desempenho dos quatro algoritmos na predição de ordens de solos, em escala de detalhe, na Serra Gaúcha, Rio Grande do Sul, Brasil. O objetivo secundário é avaliar a sensibilidade dos algoritmos à densidade de amostragem.

2. REVISÃO BIBLIOGRÁFICA

Esta revisão bibliográfica apresenta uma resenha a respeito dos conceitos fundamentais que embasam os levantamentos de solos e faz uma breve caracterização dos levantamentos convencionais e do mapeamento digital de solos. O texto aborda ainda, os métodos e técnicas utilizadas em diferentes abordagens para mapeamento digital de solos e descreve os principais atributos do terreno empregados para avaliar quantitativamente as relações solo-paisagem com vistas a estimar classes ou propriedades dos solos. Por último, aborda-se a necessidade de mensurar a qualidade dos mapas digitais estimados, através de uma avaliação de acurácia, e descrevem-se os principais indicadores utilizados.

2.1. Levantamentos de solos

Adquirir conhecimento sobre o solo tornou-se importante para a humanidade desde o início da agricultura. Civilizações surgiram e desapareceram em razão das propriedades favoráveis ou desfavoráveis dos solos. Ao longo do tempo, além da produção de alimentos, produção de energia, conservação de habitats, recursos hídricos, avaliações ambientais e inúmeras outras aplicações passaram a demandar informações sobre o recurso solo. Por fim, entender a distribuição espacial e as características dos solos tornou-se crítico para compreender o complexo balanço de processos físicos e químicos que tornam possível a vida na Terra (Scull et al., 2003).

A base conceitual para entender a gênese e a distribuição dos solos foi formalizada por Hans Jenny em 1941, a partir das idéias sobre a formação dos solos lançadas por Dokuchaev em 1883. Para Dokuchaev o solo era o resultado da interação complexa de uma série de fatores pedogenéticos. A equação proposta por Jenny (1941) relacionou os principais fatores identificados por Dokuchaev, postulando que o solo é função do clima, dos organismos e do relevo agindo sobre o material de origem durante um determinado período de tempo, conforme representado na Equação 1 abaixo:

$$S = f(cl, o, r, p, t...) \quad (1)$$

onde S é o solo, cl é o clima, o representa os organismos, r é o relevo, p é o material de origem, t é o tempo e as reticências denotam a possibilidade de incluir outros fatores eventualmente não considerados. A

teoria dos fatores de formação implica em que, se a distribuição espacial dos fatores de formação for conhecida, o solo ou suas propriedades podem ser inferidos a partir deles. Além disso, sendo o solo uma função dos fatores de formação, as relações podem ser quantificadas. Dessa forma, tal teoria serve de paradigma para os levantamentos de solos até os dias de hoje (Hudson, 1992; Scull et al., 2003; Hash, 2008).

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. Os métodos empregados nos levantamentos baseiam-se no reconhecimento de propriedades dos solos em relação à paisagem ou a variáveis que a caracterizam ou descrevem. Assim, Bui (2004) considera que um levantamento de solos consiste em um “sistema de conhecimento” e que os mapas de solos são de fato representações de um conhecimento estruturado sobre a distribuição dos solos na paisagem.

2.2. Mapeamento convencional de solos

Os levantamentos convencionais de solos contemplam um estudo do terreno e das características de perfis de solo, compreendendo descrição morfológica, caracterização física e química, classificação taxonômica e espacialização de ocorrência dos solos (mapeamento). As unidades básicas de classificação (unidades taxonômicas) são estabelecidas mediante a interpretação de dados analíticos e morfológicos de perfis representativos da menor unidade tridimensional que pode ser chamada de solo, o *pedon*. Os *pedons* com características semelhantes compõem unidades maiores, os *polipedons*, que por sua vez constituem isoladamente, ou em grupos, as unidades básicas utilizadas para compor as unidades de mapeamento (Flores et al., 2006b).

O processo de elaboração do mapa em um levantamento convencional de solos envolve basicamente três etapas. Primeiro, os dados disponíveis (fotografias aéreas, geologia, vegetação, etc.) são estudados para definir unidades de paisagem a partir das quais locais são escolhidos e visitados no campo para observar e descrever os perfis de solo. Em seguida, a partir da interpretação dos dados de campo, um modelo conceitual de solo-paisagem é desenvolvido para inferir as variações espaciais do solo. Por

último, o modelo conceitual é aplicado para inferir a distribuição espacial dos solos no restante da área e assim delimitar as unidades de mapeamento (Hudson, 1992; Ippolit et al., 2003; Elnaggar, 2007; Hash, 2008). Ao definir as fronteiras das unidades a partir desse modelo conceitual, o pedólogo usa seu conhecimento sobre os fatores de formação do solo e sobre como as variações nesses fatores ao longo da paisagem definem onde um certo tipo de solo termina e outro começa (Hudson, 1992; Hash, 2008).

A aplicação dos métodos convencionais para levantamentos de informações de solos tem uma história de várias décadas. Nos Estados Unidos, por exemplo, relatórios de levantamentos de campo e mapeamentos de solos começaram a ser produzidos pelo Departamento de Agricultura em 1899. O *US National Soil Survey Handbook* identifica padrões e especificações que são revisados regularmente e desde o início da década de 1950 constituem um protocolo para conduzir todas as fases de levantamentos de solos pelo *US National Cooperative Soil Survey* (Hempel, 2008). No Brasil, a criação da Comissão de Solos do Centro Nacional de Ensino e Pesquisas Agronômicas do Ministério da Agricultura, em 1947, deu grande impulso aos levantamentos de solos (Crivelenti, 2009). Desde então, os critérios e especificações para levantamentos em diferentes escalas vêm sendo aperfeiçoados continuamente e mais recentemente foram reunidos em publicações de referência (Embrapa, 2006; IBGE, 2007).

Atualmente os mapas e relatórios produzidos por levantamentos convencionais são a forma mais comum pela qual a informação sobre esse recurso natural complexo tem sido disponibilizada para a comunidade. Em muitos casos os mapas convencionais constituem o único registro disponível da natureza altamente variável das relações solo-paisagem (Scull, et al., 2005; Hempel et al., 2008). Entretanto, não obstante representem o conhecimento acumulado pela Ciência do Solo durante décadas, os levantamentos convencionais tem recebido críticas com relação às suas limitações para atender as demandas crescentes da sociedade sobre diferentes informações de solos.

Entre as limitações mais citadas destaca-se a indisponibilidade generalizada de mapas de solos. Segundo Gonzáles et al. (2007) e Sanchez et al. (2009), cerca de 109 países possuem mapas convencionais de solos em

escalas 1:1.000.000 ou maiores, mas eles cobrem somente 31% da superfície da Terra emersa e livre de gelo. O restante dispõe apenas do mapa de solos na escala 1:5.000.000 elaborado pela FAO. O tempo e o custo necessários para a execução das diferentes fases do levantamento restringem severamente a expansão e atualização dos mapeamentos e provavelmente sejam os maiores responsáveis pela lacuna de informações de solos (McBratney et al., 2003). O intervalo de tempo que passa desde os primeiros trabalhos de campo até a conclusão e publicação do mapeamento pode chegar a vários anos. Além da demora em si, esse intervalo também afeta a consistência do mapa final de solos, pois dificulta revisões e eventuais correções pelos autores ao longo do processo (Weber et al., 2008).

Além disso, o processo convencional depende de avaliações mentais qualitativas das propriedades dos solos em relação aos fatores de formação, as quais raramente são documentadas. Essas avaliações são altamente dependentes da experiência do pedólogo, tornando improvável a reprodução ou atualização de um levantamento com os mesmos critérios, além de dificultar a transferência do conhecimento para outro profissional que venha a substituí-lo na tarefa (Scull et al., 2003; Hash, 2008). Os mapas convencionais também são de difícil compreensão por usuários fora da Ciência do Solo, porque poucos estão familiarizados com os conceitos e terminologias adotadas. Apesar de tentativas de traduzir o conteúdo na forma de legendas interpretadas (classes de fertilidade, profundidade, aptidão, etc.), tais mapas ainda são uma forma de comunicação pouco eficiente para a maioria dos usuários potenciais (Bui, 2004).

A falta de informações sobre a qualidade é mais um aspecto criticado, já que restringe o uso dos mapas convencionais disponíveis. Um mapa convencional é uma simplificação da distribuição complexa e contínua dos solos na paisagem, onde estes são representados como unidades homogêneas em polígonos discretos, de fronteiras definidas. Essa simplificação resulta tanto da necessidade de usar uma representação bidimensional de fronteiras rígidas quanto das restrições impostas pela escala do mapeamento (Heuvelink & Webster 2001; Shi et al., 2004). Por maior que seja o conhecimento e a dedicação do pedólogo, as relações solo-paisagem reais serão sempre mais complexas do que é possível registrar em um mapa

convencional (Hempel et al., 2008). Assim, dispor de uma medida sobre o quanto o mapa se aproxima da realidade é fundamental para avaliar sua adequação a um determinado propósito.

As limitações mencionadas estão mais relacionadas aos métodos e procedimentos de execução dos levantamentos convencionais do que ao conhecimento da pedologia que lhe serve de base. Eles ainda fazem pouco ou nenhum uso de tecnologias hoje disponíveis para coleta e análise de dados espaciais, como SIG, GPS e imagens de sensoriamento remoto (Scull et al., 2003). Com o apoio desses recursos, as relações solo-paisagem, avaliadas qualitativamente em levantamentos convencionais, podem ser analisadas de forma quantitativa e, portanto, mais objetiva. Para Sanchez et al. (2009) uma nova geração de pedólogos necessita ser treinada nessa abordagem com vistas a disponibilizar informações de solos na velocidade e com a qualidade necessária às demandas da sociedade.

2.3. Mapeamento digital de solos

Segundo Lagacherie & McBratney (2007), o mapeamento digital de solos consiste na criação e alimentação de sistemas espaciais de informação de solos, através do uso de modelos numéricos para a inferência das variações espaciais e temporais dos tipos de solos ou de suas propriedades, a partir de observações e conhecimento dos solos e de variáveis ambientais correlacionadas.

O mapeamento digital de solos começou na década de 1970 e acelerou-se significativamente nos anos 1980, com o advento de novas tecnologias de coleta, processamento e análise de informações espaciais, como os Sistemas de Informação Geográfica (SIG), o Sistema de Posicionamento Global (GPS) e o sensoriamento remoto. A partir do início da década de 1990, avanços sensíveis na capacidade de processamento dos computadores e em softwares de estatística e modelagem deram grande impulso aos estudos de mapeamento digital de solos. Mais recentemente, o acesso a informações através de recursos da rede mundial de computadores possibilitou o intercâmbio de experiências e a busca de consensos para a implementação de sistemas de mapeamento digital em nível operacional (McBratney et al., 2003; Rossiter, 2005; Sanchez et al., 2009).

A rapidez e facilidade de processamento de informações geográficas proporcionada pelas novas tecnologias e o número crescente de estudos voltados à predição de propriedades ou classes de solos resultou ainda no advento e crescimento da pedometria como ramo da Ciência do Solo. A pedometria procura estabelecer relações matemáticas e (geo)estatísticas entre classes ou propriedades do solo e variáveis quantitativas utilizadas como preditoras daquelas. McBratney et al. (2000) definem pedometria como a aplicação de métodos matemáticos e estatísticos para o estudo da distribuição e gênese de solos.

Entre as diferentes abordagens em mapeamento digital de solos destaca-se o modelo *s.c.o.r.p.a.n.* proposto por McBratney et al. (2003), representado na Equação 2 abaixo:

$$S_{c,p} = f(s.c.o.r.p.a.n.) \quad (2)$$

onde:

$S_{c,p}$ = solo - classe ou propriedade;

s = solo ou propriedades do solo em um dado ponto;

c = clima, propriedades climáticas do ambiente em um dado ponto;

o = organismos, incluindo cobertura vegetal e uso do solo;

r = relevo, incluindo atributos do terreno e classes;

p = material de origem, litologia;

a = idade, tempo;

n = espaço, posição espacial, georreferenciamento.

Adaptado da equação clássica de Jenny (1941), o modelo *s.c.o.r.p.a.n.* propõe explicar, por diversos métodos, as relações entre os fatores de formação do solo, bem como predizer de forma quantitativa as relações entre classes ou propriedades dos solos e o ambiente, espacialmente. A inclusão da posição espacial (n) elevou as geotecnologias à condição de ferramentas indispensáveis para o mapeamento digital de solos.

Apesar dos avanços e embora várias abordagens apóiem-se na teoria dos fatores de formação, o mapeamento digital de solos ainda é visto com ceticismo por muitos pedólogos. Uma das ressalvas feitas por estes é a ausência de um padrão nos trabalhos de mapeamento digital, enquanto os levantamentos convencionais dispõem de protocolos há quase um século. Todavia, ao analisar a rotina dos mapeamentos digitais constata-se que a maioria segue algumas etapas que, na essência, podem ser consideradas

análogas àquelas usadas nos levantamentos convencionais. Assim como nestes, o mapeamento digital faz uso de informações coletadas ou disponíveis em pontos de observação de solos. Essas informações são utilizadas para ajustar um modelo quantitativo com variáveis relativas às condições do ambiente nos mesmos locais e o modelo ajustado é depois empregado para prever propriedades do solo ou classes de solos para o restante da área (Lagacherie, 2008).

MacMillan (2008) detalha um pouco mais essas etapas e afirma que, independente da abordagem, dos objetivos e dos métodos empregados, o processo de mapeamento digital de solos envolve basicamente seis passos. O primeiro passo consiste em definir o objetivo do mapeamento: prever classes de solos ou propriedades individuais de solos. O segundo passo é identificar e obter dados de entrada relevantes para se alcançar o objetivo proposto, incluindo pontos com informação de referência sobre classes ou propriedades de solos e variáveis que podem ser usadas como preditoras, tais como atributos do terreno derivados de MNT, imagens de sensoriamento remoto e mapas existentes. O terceiro passo envolve o desenvolvimento de equações para prever propriedades dos solos ou de regras de classificação para estimar classes de solos. O quarto passo consiste em aplicar as equações ou as regras de classificação previamente desenvolvidas e o quinto corresponde à avaliação do sucesso ou acurácia da predição das propriedades ou classes, comparando-se o resultado com dados de referência coletados em pontos de observação. Dependendo do resultado, pode-se retornar às etapas anteriores para revisar ou adequar objetivos, variáveis ou as próprias equações. Uma vez alcançado um grau de acurácia aceitável, o sexto e último passo, que é a geração dos mapas finais, pode ser finalmente executado.

Ao comparar essas etapas com as dos levantamentos convencionais verifica-se que o mapeamento digital de solos não representa mais do que uma evolução da pedologia convencional, com base no conhecimento acumulado pela Ciência do Solo e nas tecnologias disponíveis. A maior diferença é que, no lugar de um modelo conceitual qualitativo sobre a distribuição dos solos na paisagem, são empregados modelos quantitativos. Uma vez estabelecido um modelo quantitativo a partir das variáveis selecionadas, ele é aplicado uniformemente sobre toda a área a ser mapeada, em uma análise *pixel a pixel*,

possibilitando a obtenção de mapas espacialmente mais detalhados que os convencionais (Zhu, 2000; Hempel et al., 2008; MacMillan, 2008). Adicionalmente, os resultados obtidos podem ser submetidos a uma avaliação de acurácia também quantitativa, gerando assim uma medida sobre a qualidade do mapa produzido, o que raramente é feito em levantamentos convencionais.

A emergência do mapeamento digital, todavia, não significa que os conceitos convencionais de levantamentos de solos devem ser abandonados. Ao contrário, convém que mapeadores digitais de solos trabalhem em parceria com pedólogos experientes com vistas a desenvolver metodologias robustas e operacionais que possibilitem atender as demandas por informações sobre solos. A predição de classes ou propriedades de solos necessita de um bom conjunto de informações coletadas em pontos de observação para ajustar os modelos e para avaliar a qualidade dos resultados obtidos. Assim, os motivos para cooperação tornam-se óbvios: de um lado existe um grande volume de dados de perfis de solo coletados em levantamentos já existentes e de outro há um extenso conjunto de técnicas pedométricas que necessitam de observações de solos para ajustar os modelos. Combinar as duas pode ser uma forma de obter rapidamente novos mapas digitais mais detalhados, sem refazer levantamentos caros e extensivos (McBratney et al., 2003; Hengl et al., 2007; Weber et al., 2008).

Embora propostas tenham sido formuladas (Zhu, 1997; Shi et al., 2004; Gray et al., 2009), ainda não há procedimentos, padrões ou protocolos estabelecidos e aceitos para a geração de mapas digitais de solos (Hempel et al., 2008). De fato é muito difícil estabelecer protocolos porque as ferramentas disponíveis permitem desenvolver tantos modelos quantos forem as variáveis à disposição e os tipos de resultados desejados. Obviamente um dos objetivos da investigação científica nessa área é a obtenção de modelos explicitamente fundamentados nos fatores de formação do solo, que possam ser validados e extrapolados para regiões fora daquela onde foram criados (Walter et al., 2007). Entretanto, a falta de protocolos específicos não representa necessariamente um obstáculo para o mapeamento digital, já que a acurácia do resultado pode ser mensurada. Neste caso, enquanto modelos gerais não são estabelecidos, é mais eficaz garantir a realização de uma avaliação de

acurácia a fim de que o usuário final disponha de uma medida de qualidade, do que tentar restringir o processo de elaboração de um mapa digital.

2.3.1. Principais métodos e técnicas

As diferentes metodologias de modelagem empregadas para avaliar e quantificar as relações entre as variáveis preditoras selecionadas e as propriedades ou classes de solos a serem estimadas incluem o uso de um grande número de técnicas de processamento e análise de dados, entre as quais regressões múltiplas, geoestatística (*kriging*), lógica de conjuntos difusos (lógica *Fuzzy*), redes neurais artificiais e árvores de decisão. A gama de recursos e ferramentas utilizada inclui bases cartográficas, mapas existentes (solos, geologia, vegetação, etc.), MNT, bancos de dados, produtos de sensoriamento remoto (dados de sensores, imagens de satélite, fotografias aéreas), receptores GPS (*Global Positioning System*), *softwares* de Sistemas de Informação Geográfica (SIG), de processamento de imagens e de análises estatísticas, entre outras.

2.3.1.1. Regressões múltiplas

O uso de regressões lineares provavelmente seja o método mais disseminado no meio científico para estudar a relação entre variáveis ou fazer previsões. As regressões lineares múltiplas assumem a existência de uma relação linear entre a variável dependente e duas ou mais variáveis independentes. Por exemplo, no caso de três variáveis independentes a regressão linear múltipla pode ser escrita como a Equação 3 abaixo:

$$Y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 \quad (3)$$

onde Y é a variável dependente (ou variável resposta), x₁, x₂ e x₃ são as variáveis independentes, a é o intercepto e b₁, b₂ e b₃ são os coeficientes angulares das variáveis independentes (x₁, x₂, e x₃, respectivamente). O intercepto representa o valor de Y quando os valores das variáveis independentes for zero, e o coeficiente indica a taxa de alteração de Y para cada unidade de incremento na variável independente correspondente (Bailey et al., 2003).

Em geral a equação de regressão linear múltipla é calculada através de um método de mínimos quadrados e o grau de associação entre a variável dependente e as variáveis independentes é expresso pelo coeficiente de

determinação (R^2), cuja significância estatística é avaliada com base no teste F ou teste t . Alguns pressupostos básicos devem ser observados no uso de regressões lineares:

- a) A variável dependente deve obedecer a uma distribuição normal;
- b) O número de observações precisa ser maior que o número de variáveis independentes;
- c) Não deve existir relação linear entre as variáveis independentes.

Um caso especial de regressões múltiplas são as regressões logísticas, nas quais a variável resposta é discreta, como por exemplo, a classe de solo. As variáveis independentes podem ser tanto discretas quanto contínuas. As regressões logísticas múltiplas são uma adaptação do modelo original desenvolvido para uma variável independente binária. Além de a variável dependente ser do tipo discreta, as regressões logísticas se diferenciam das lineares pela sua função de ajuste, que segue a curva logística (sigmóide).

As regressões logísticas têm a desvantagem de não possuírem um valor de r^2 como as regressões comuns, necessitando-se de estratégias mais complexas para excluir variáveis preditoras desnecessárias do modelo, bem como para verificar a adequação do modelo quanto à estimativa da resposta (Wong & Mason, 1985). Em geral a verificação é feita pelo teste dos coeficientes angulares e a validação é feita com um conjunto de novos dados ou, na falta destes, com uma sub-amostra dos dados usados para estabelecer a curva. Outra forma de avaliar a acurácia de mapas produzidos com regressões múltiplas é utilizar índices de concordância, como o índice *Kappa* (Rossiter, 2004a).

Em relação a regressões logísticas múltiplas e classes de solo, Bailey et al. (2003) efetuaram a estimativa de classes de solos na Inglaterra usando um mapa de solos pré-existente na escala 1:250.000 e um MNT. A correspondência entre os valores estimados e a verdade de campo variou de 47% a 65% em três formas de relevo: planalto, vale e escarpa. As conclusões apontam que as regressões logísticas múltiplas são um método válido para estimativa de classes de solo e para aumentar o detalhamento de mapas de solos pré-existentes.

Ziadat (2005) empregou regressão logística múltipla para avaliar o uso de atributos do terreno derivados de um MNT como preditores de alguns atributos do solo. As conclusões apontam que o método é promissor e, quando a área for extensa, a precisão pode ser melhorada pela subdivisão em áreas menores e aplicando regressões ajustadas para cada subdivisão. Além disso, o uso de regressões permitiu a predição dos atributos avaliados com uma resolução melhor do que os mapas de solos convencionais.

Giasson et al. (2006) empregaram regressões logísticas múltiplas na predição de ocorrência de classes de solos a partir de áreas de referência, utilizando um mapa de solos pré-existente na escala 1:50.000 e parâmetros do terreno derivados de um MNT. Os parâmetros do terreno como elevação, distância dos rios, acúmulo de fluxo e índice de umidade topográfica, foram as variáveis que melhor explicaram a distribuição das classes de solos e a simplificação da legenda aumentou a precisão do método na predição da distribuição dos solos.

Giasson et al. (2008) também usaram regressões logísticas para estimar classes de solos para diversas áreas no sul do Brasil, empregando mapas de solos existentes nas escalas 1:30.000 a 1:50.000 e variáveis do terreno derivadas de um MNT. Avaliando a capacidade dos modelos em reproduzir os mapas originais, os autores encontraram índice *Kappa* de 0,36 para a legenda original e de 0,54 para uma legenda simplificada, com agrupamento das classes. Os baixos valores de acurácia na estimativa das classes individuais foram atribuídos à distribuição espacial complexa das mesmas e à grande variabilidade do material de origem da região, para o qual não há informações de qualidade disponíveis.

2.3.1.2. Geoestatística

A geoestatística começou na indústria da mineração no início da década de 1950, como uma tentativa de melhorar a estimativa de ocorrência de depósitos de ouro. A idéia original foi desenvolvida pelo engenheiro de minas D. G. Krige e pelo estatístico H. S. Sichel e publicada em 1951. O objetivo básico era estimar as propriedades do terreno em toda a superfície de uma área de interesse com base em um conjunto de pontos amostrais distribuídos na mesma. O pressuposto era de que, quanto mais próximo de um ponto amostrado estivesse o local a ser estimado, maior era a probabilidade de suas

características serem parecidas a esse ponto. Tendo-se um conjunto de pontos, era possível ponderar a influência de cada um sobre o local a ser estimado.

Quase uma década depois da proposição inicial, o matemático francês G. Matheron aprimorou as fórmulas originais e estabeleceu o campo conhecido como geoestatística, uma área da estatística especializada em análise e interpretação de dados geograficamente referenciados. Entretanto, até hoje o termo *Kriging* é usado como sinônimo de geoestatística (Hengl, 2007).

Em um contexto mais pragmático, a geoestatística é uma ferramenta analítica para dados obtidos em pontos de observação e para estimar mapas em toda a área de interesse a partir dos valores de variáveis amostradas em tais pontos, tanto com interpolação quanto com extrapolação (Hengl, 2007). O mapeamento geoestatístico pode ser definido como a produção analítica de mapas utilizando observações de campo, informações auxiliares e um programa de computador que calcula valores nas localizações de interesse.

Segundo Hengl (2007), existem na literatura muitas definições e termos distintos para a geoestatística, vários deles criados simplesmente para designar uma aplicação específica das técnicas, mas todos se baseiam nos pressupostos básicos do método *kriging*. A versão padrão da estimativa *Kriging* é chamada de *Kriging* ordinário ou *Kriging* simples, onde as previsões são baseadas no modelo da Equação 4 a seguir:

$$Z(s) = \mu + \varepsilon'(s) \quad (4)$$

onde μ é a função constante estacionária (média global) e $\varepsilon'(s)$ é a parte estocástica do variograma espacialmente correlacionada. O variograma descreve o comportamento das diferenças (na forma de semivariância) entre os pontos amostrais em função do afastamento entre os mesmos, permitindo avaliar a dependência espacial (efeito de autocorrelação espacial). O esperado é que as diferenças sejam pequenas entre pontos próximos e aumentem conforme aumenta a distância entre os pontos, até se estabilizar (Figura 1). A partir do variograma é possível determinar a importância que deve ser dada a cada ponto amostral no momento de estimar a variável alvo para um local não amostrado, definindo os pesos em função da distância.

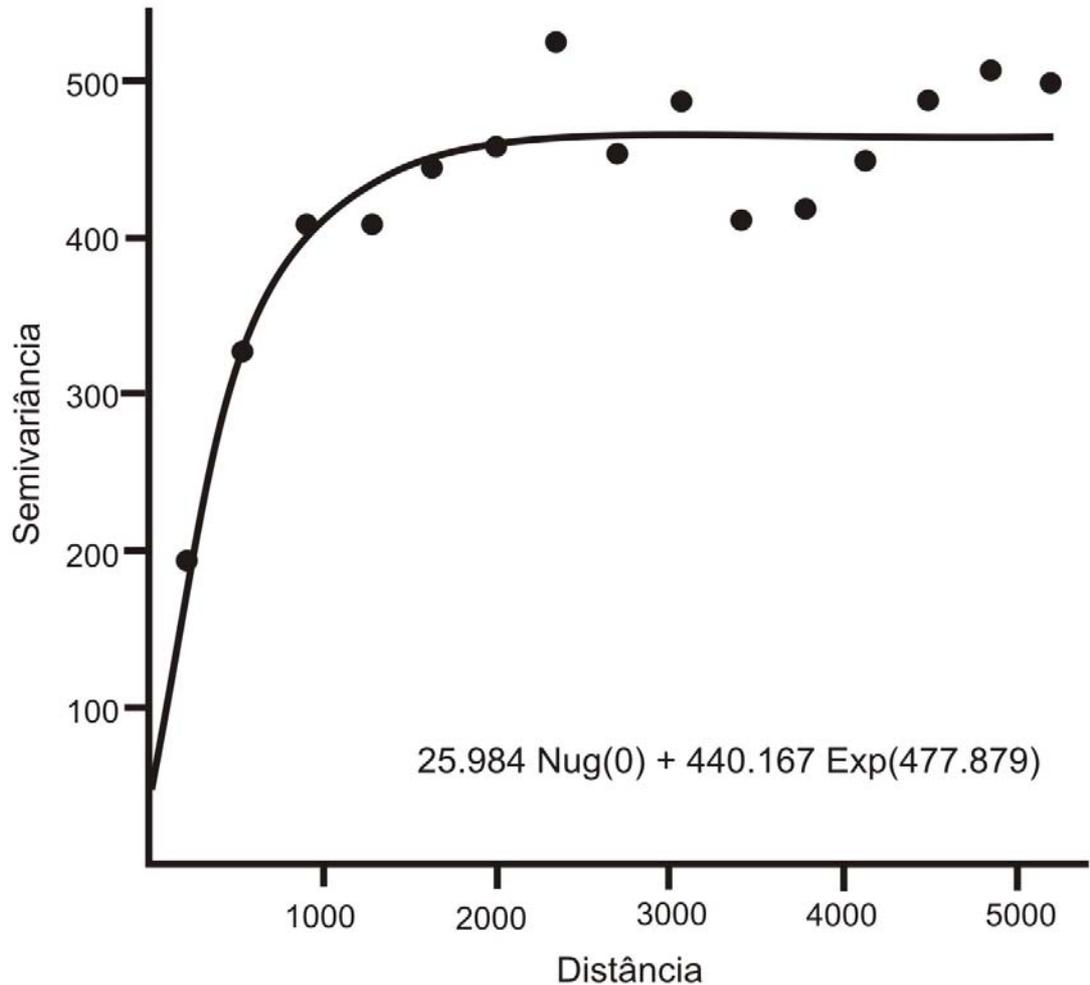


Figura 1. Exemplo de um variograma (Hengl, 2007).

Quando outras variáveis correlacionadas com a variável alvo (correlação espacial cruzada) estiverem disponíveis em densidade suficiente para a área de interesse, elas podem ser usadas para prever o valor de uma variável de solo em locais não amostrados. Para tanto, a relação entre os preditores e a variável alvo é primeiro modelada e depois o modelo é aplicado aos locais não amostrados, usando o valor conhecido das variáveis auxiliares nesses locais. Entre as variáveis auxiliares mais comuns para mapeamento digital de solos encontram-se os atributos do terreno derivados de MNT, imagens de sensoriamento remoto e dados de solos ou geologia coletados em pontos de observação ou em mapas existentes. Essa abordagem da geoestatística é definida como *regressão kriging*, *correlação kriging* ou *co-kriging* (Figura 2).

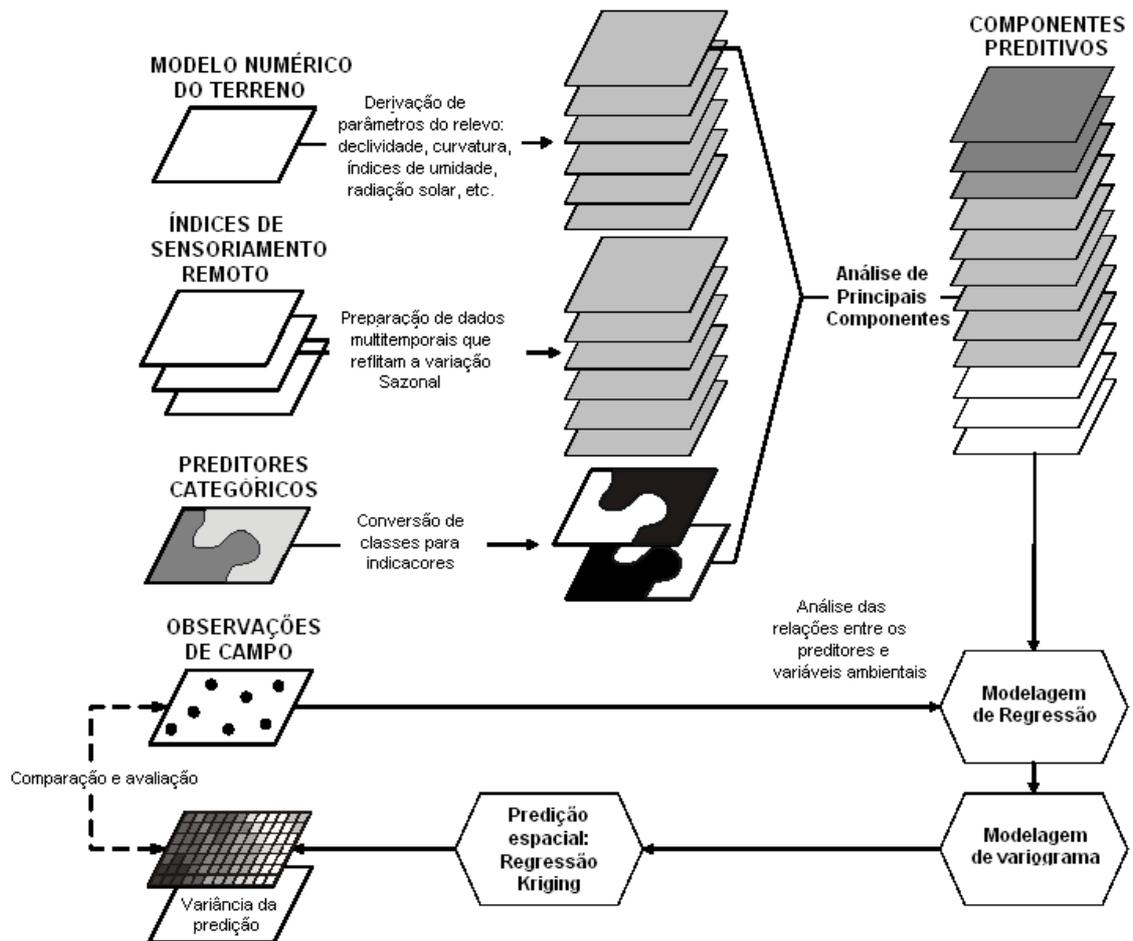


Figura 2. Esquema de mapeamento digital de solos com *regressão-kriging* (Dobos et al., 2006).

A geoestatística tem sido uma abordagem bastante utilizada em trabalhos de mapeamento digital de propriedades de solos ou de classes de solos, tanto para a interpolação de variáveis de entrada quanto nas estimativas de saída propriamente ditas. De acordo com Mendonça-Santos et al. (2000), entre as técnicas que permitem o uso de modelos contínuos, a geoestatística é vista como a mais apropriada para uso na Ciência do Solo.

Bourennane et al. (2000), por exemplo, compararam o desempenho de interpolação geoestatística com variáveis auxiliares (*regressão kriging*) e de regressão linear múltipla para prever a espessura do solo, empregando diferentes densidades de amostragem. Os resultados mostraram que as duas técnicas foram igualmente capazes de prever a variável de interesse, mas a geoestatística apresentou estimativas melhores com o aumento da densidade de amostragem enquanto a regressão linear múltipla manteve-se estável com todas as densidades testadas.

Hengl et al. (2007) testaram o uso de classificação supervisionada por máxima verossimilhança, regressão logística múltipla, *regressão kriging* e classificação de distância taxonômica para prever a distribuição de classes de solos no Irã, com base em parâmetros do terreno e em índices de sensoriamento remoto e 4.250 perfis de solos. A acurácia foi avaliada através do índice *Kappa*, sendo que o método *regressão kriging* apresentou os melhores resultados.

2.3.1.3. Lógica de conjuntos difusos (lógica fuzzy)

Conjuntos difusos são conjuntos (ou classes) sem fronteiras definidas, onde a transição entre membros e não membros de uma posição no conjunto é gradual, representada por uma função de pertinência. Um conjunto difuso é caracterizado por um grau de pertinência que varia de 0,0 a 1,0, indicando um incremento contínuo desde a não pertinência até a pertinência completa ao conjunto (Figura 3). Dessa forma, a lógica de conjuntos difusos é capaz de incorporar a incerteza presente em diferentes tipos de dados, ao contrário da lógica clássica de conjuntos *Booleanos* que só tem duas possibilidades: não (0) ou sim (1).

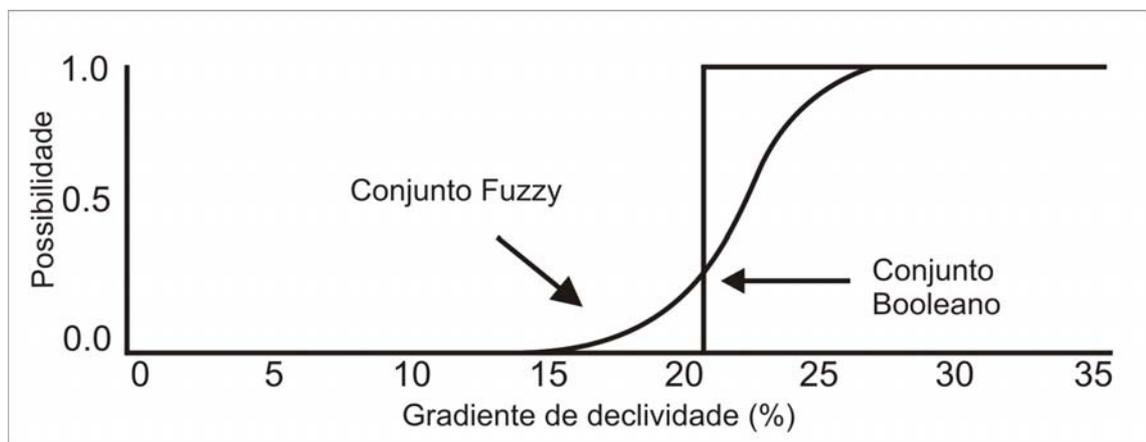


Figura 3. Exemplo de função de pertinência a conjunto difuso em contraste com uma função de conjunto booleano.

Incerteza, imprecisão e ambigüidade são inevitáveis e inerentes em sistemas naturais como o solo. Aplicações da teoria de conjuntos difusos na Ciência do Solo permitem lidar com as incertezas, principalmente devido às bordas imprecisas entre diferentes classes (Minasny & McBratney, 2002). Segundo McBratney et al. (2003), a teoria de conjuntos difusos oferece uma rica base matemática para entender problemas de decisão e para construir

regras de decisão na avaliação e combinação de critérios para o mapeamento digital de solos. O desenvolvimento de técnicas de mapeamento digital baseadas em lógica de conjuntos difusos tem atraído atenção devido à sua habilidade em capturar e representar a natureza contínua da variação espacial do solo. Em função disso, abordagens para mapeamento digital de propriedades ou de classes de solos baseadas em lógica difusa têm sido propostas com frequência, tornando-a uma das técnicas mais investigadas (Zhu et al., 1996; Zhu et al., 2001).

Há basicamente duas abordagens de conjuntos difusos em mapeamento digital de solos. A primeira procura particionar as observações feitas sobre um espaço multivariado em grupos naturais relativamente estáveis, ou seja, em grupos formados por similaridade de características. A segunda consiste em utilizar conjuntos difusos para classificação de solos, onde uma função de pertinência é definida *a priori*, com base no conhecimento especialista ou em definições convencionalmente impostas, para então alocar os dados conforme o grau com que eles se ajustam aos requerimentos das classes (McBratney et al., 2000; Qi & Zhu, 2003). Em ambos os casos a variação espacial do solo é expressa como variação espacial de pertinência a classes de solos, a qual é usada para produzir mapas de classes de solos ou para prever a variação espacial de propriedades específicas dos solos (Zhu et al., 1996; Zhu, 1997; Qi et al., 2006).

Shi et al. (2004), por exemplo, usaram lógica de conjuntos difusos para identificar classes de solos e para estimar algumas propriedades dos solos em uma bacia hidrográfica de 5 km² em Pleasant Valley, sudoeste do Estado de Wisconsin, nos Estados Unidos. A metodologia envolveu uma etapa de captura do conhecimento sobre as relações solo-paisagem, envolvendo a participação de pedólogos com experiência de campo na área de estudo, e outra etapa para a predição de classes e de propriedades a partir das relações previamente estabelecidas. Os autores obtiveram concordância de 86% na identificação das classes pelo mapa estimado, contra 46% do mapa convencional, usando dados de referência observados em 57 pontos. Além da melhor exatidão, ainda foi possível obter mapas com a variação espacial da textura do horizonte A e da profundidade do horizonte C dentro das classes, as quais são consideradas homogêneas no mapa convencional.

Zhu et al. (2008) também usaram um procedimento de classificação por lógica de conjuntos difusos para identificar combinações únicas de fatores ambientais associadas a classes típicas de solos. O objetivo foi avaliar formas de racionalizar a coleta de dados de campo para desenvolver modelos de solo-paisagem para mapeamento digital, de forma a conduzir futuras amostragens em locais de classes mais puras. A área de estudo foi uma bacia hidrográfica de 60 km² na província de Heilongjiang, na China. Os autores encontraram exatidão geral de 76% entre as classes estimadas através do modelo solo-paisagem e as classes observadas em campo, resultado que consideraram promissor para o mapeamento digital de solos em áreas com disponibilidade limitada de dados.

2.3.1.4. Redes neurais artificiais

As redes neurais artificiais são algoritmos de aprendizagem de máquina, ou aprendizagem indutiva, que tentam simular a forma como o cérebro humano processa informações. Elas são não lineares e podem ser entendidas como funções matemáticas complexas que convertem dados de entrada (por exemplo, variáveis de relevo) para uma saída desejada, geralmente do tipo categórica (por exemplo, classes de solos).

A estrutura de uma rede neural típica contém uma camada de entrada, uma ou mais camadas intermediárias, ocultas, e uma camada de saída (Figura 4). Cada camada contém um conjunto de nós (ou neurônios) que estão conectados a todos os nós das outras camadas, com a possibilidade de atribuir diferentes pesos em cada conexão (Richards & Jia, 1999). Existem muitas variantes de redes neurais artificiais com relação à abordagem usada para o treinamento, aos métodos de atribuição dos pesos dos neurônios e à conectividade dos neurônios dentro de uma camada e entre as diferentes camadas.

Uma rede neural artificial resolve um problema primeiro desenvolvendo uma memória através da associação de um grande número de padrões de entrada com um conjunto limitado de saída, utilizando exemplos, e depois aplicando essa memória para produzir uma saída quando um determinado padrão de entrada é fornecido (Qi & Zhu, 2003). Da mesma forma que acontece com o cérebro humano, redes neurais artificiais precisam ser treinadas para que aprendam a executar o processo desejado. O treinamento

exige um conjunto de dados amostrais, e pode ser usado para configurar a rede para o reconhecimento de padrões ou para o agrupamento dos dados em classes (Minasny & McBratney, 2002).

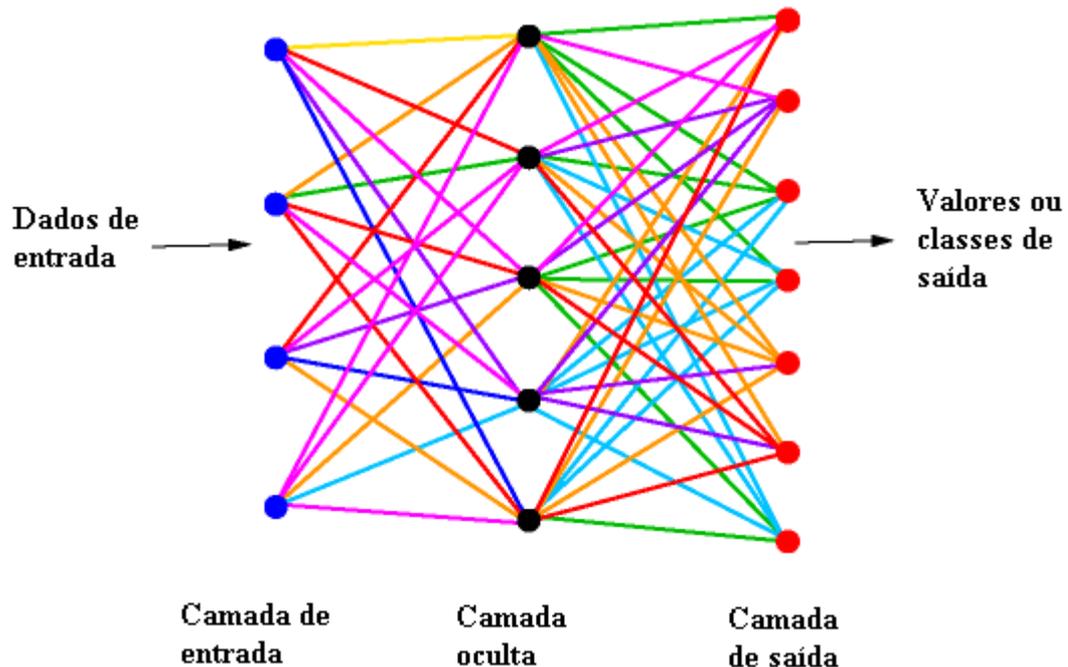


Figura 4. Esquema de uma rede neural com uma camada oculta (diferentes cores das linhas indicam conexões com pesos diferentes entre nós).

O principal propósito do treinamento consiste em obter os pesos apropriados para as conexões entre a camada de entrada e a(s) camada(s) oculta(s) e entre esta(s) e a camada de saída, de forma análoga ao que ocorre com as conexões sinápticas do cérebro. Uma vez estabelecidos os pesos das conexões, são eles que possibilitam à rede neural extrair informações úteis das variáveis de entrada para fazer a predição da variável de saída, ou variável resposta, como classes de solos, por exemplo (Richards & Jia, 1999).

Em função da concepção, a forma de operação de uma rede neural não possibilita ao usuário controle sobre as funções que produzirão o resultado, o que rende ao método algumas críticas pelo caráter de “caixa preta”. O controle é exercido por alterações na estrutura da rede, pela definição das variáveis preditoras usadas como entrada e pelo tamanho do conjunto de amostras de treinamento. Poucas amostras podem não representar adequadamente os padrões ou categorias em análise, enquanto amostras em

demasia podem causar sobreposição. Por essa razão, geralmente os algoritmos de redes neurais possuem recursos para monitorar os erros e orientar o dimensionamento da estrutura da rede e do conjunto de amostras (Richards & Jia, 1999).

Uma das vantagens das redes neurais é que sua utilização não necessita da definição de nenhum modelo prévio para descrever a distribuição espacial da variável de interesse. Redes neurais são um meio efetivo de modelar propriedades descritivas e quantitativas dos solos e podem ajudar a descobrir relações ocultas nos dados, difíceis de detectar por outros meios devido às interações complexas que ocorrem no solo (Levine et al., 1996). Elas têm sido bastante utilizadas para o processamento e classificação de imagens de sensoriamento remoto e recentemente também vem sendo aplicadas com sucesso no mapeamento digital de solos (McBratney et al., 2003).

Levine et al. (1996), por exemplo, testaram diferentes arquiteturas de redes neurais para classificar a estrutura do solo a partir de dados analíticos de 390 amostras, obtendo uma concordância de 79% no melhor resultado. Zhu (2000), empregou redes neurais na predição de classes de solos na escala 1:10.000 em uma bacia hidrográfica na Floresta Experimental de Lubrecht, uma região montanhosa no oeste do Estado de Montana, Estados Unidos. As variáveis preditoras utilizadas foram geologia, cobertura vegetal e quatro atributos do terreno derivados de um MNT: elevação, declividade, orientação e curvatura de perfil. Com base em 64 pontos observados em campo, os resultados mostraram que as classes do mapa estimado por rede neural foram concordantes em 49 pontos (exatidão geral de 77%).

Behrens et al. (2005) também usaram redes neurais para estimar classes de solos em uma área de 600 km² na região do Palatinado, sudoeste da Alemanha, a partir de um grande número de variáveis como descritoras dos fatores de formação e de um mapa existente de solos. A exatidão geral média encontrada pelos autores para os mapas estimados foi de 72%, sendo o melhor resultado obtido com o uso simultâneo de variáveis descritoras do relevo, uso do solo e geologia.

Contador et al. (2006) empregaram redes neurais e variáveis topográficas derivadas de um MNT de alta resolução para estimar a distribuição espacial da umidade superficial do solo em uma bacia hidrográfica

de 99,5 ha no sudoeste da Espanha. Os autores obtiveram estimativas com erros variando de 5,5 a 8,7% e constataram que a curvatura vertical foi a variável topográfica mais relacionada com a umidade. As redes neurais mostraram-se altamente eficazes e, combinadas com uma estratégia de amostragem adequada, permitiram monitorar a umidade com um número reduzido de amostras.

Chagas et al. (2007) testaram a predição de classes de solos na escala 1:50.000 em uma região montanhosa do Rio de Janeiro a partir de atributos do terreno e índices derivados de imagens de sensoriamento remoto, empregando uma rede neural *Multi-layer Perceptron* (MLP) e o algoritmo da Máxima Verossimilhança Gaussiana (MAXVER). Usando observações de campo para o treinamento dos algoritmos e para a avaliação de acurácia, os autores obtiveram valores de exatidão geral e índice *Kappa* respectivamente de 70,83% e 0,678 com a rede neural MLP e de 50,69% e 0,46 com o algoritmo MAXVER.

2.3.1.5. Árvores de decisão

Árvores de decisão ou de classificação, assim como redes neurais artificiais, são algoritmos de aprendizagem de máquina que estabelecem modelos através de exemplos, mas com uma forma de abordagem distinta. Elas simulam o processo de abstração humana através de uma categorização hierárquica, obtendo regras similares a uma chave de classificação. Usando um conjunto de dados amostrais como exemplo, uma árvore de decisão divide recursivamente os dados para formar subconjuntos homogêneos, resultando em uma árvore hierárquica de regras de decisão (Behrens & Scholten, 2006), que é composta de basicamente três componentes:

- a) A raiz, ponto de partida da árvore;
- b) O nó, conexão entre a raiz, os nós subseqüentes e as folhas;
- c) A folha, que representa uma classe homogênea.

Na prática, uma árvore de decisão aplica uma série de regras de decisão binárias sucessivas para enquadrar os dados, partindo do conjunto completo de dados (raiz). Cada decisão divide os dados em duas novas classes homogêneas, e cada nova classe pode ser novamente dividida em duas ou mais classes com base em outra regra. Cada subdivisão baseada em uma regra é chamada de nó, podendo ser criados tantos nós quantos forem

necessários para produzir as classes finais desejadas ou para obter subconjuntos tão homogêneos quanto possíveis (Figura 5). 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 & Franklin, 2001).

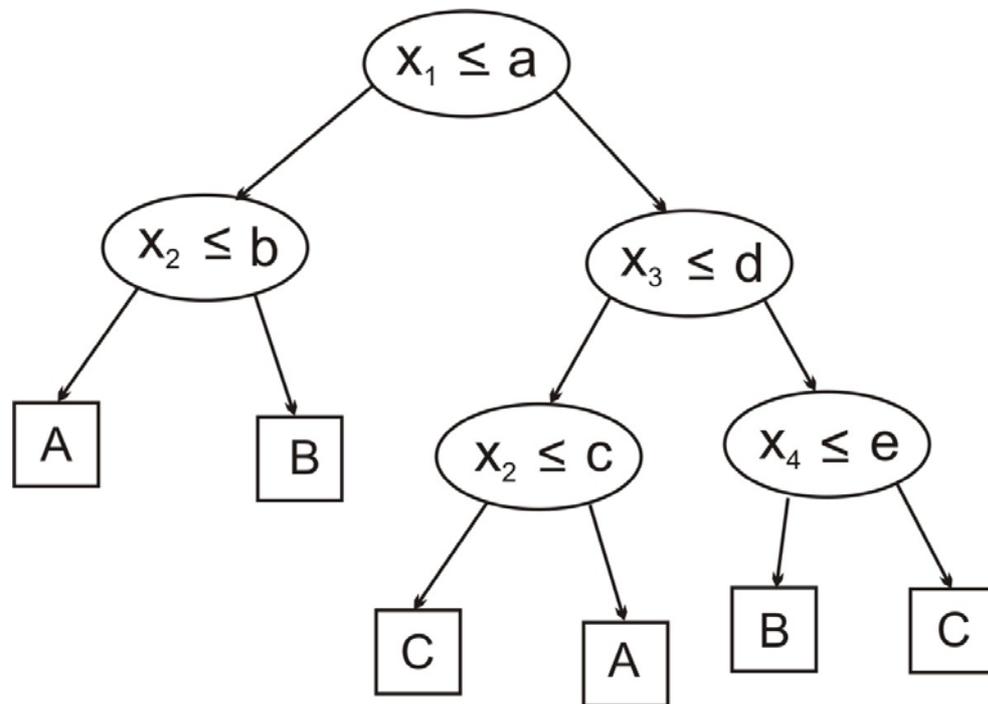


Figura 5. Esquema de uma árvore de decisão com quatro variáveis e três classes. x_i representa as variáveis; a, b, c, d, e representam os limites da regra de decisão; A, B, C representam as classes (Zhou et al., 2004).

Para evitar a geração de árvores muito complexas, com excessivo número de folhas, muitas das quais podem ter pouca extensão ou até mesmo representar um único ponto, é necessário “podar” a árvore. Esse procedimento consiste em reagrupar folhas demasiadamente pequenas e é implementado na forma de um grande número de métodos, os quais constituem a essência das diferenças entre os vários algoritmos de árvores de decisão disponíveis (Scull et al., 2005).

De acordo com essa lógica, uma árvore de decisão pode ser entendida como um modelo preditivo onde um nó corresponde a uma variável e as folhas representam o valor estimado para a variável resposta a partir dos valores das variáveis preditoras ao longo do caminho, desde a raiz até a folha. A estratégia básica é transformar uma decisão complexa em várias decisões simples. Uma das vantagens práticas desse método é que as árvores de

decisão são geralmente simples de entender e fáceis de interpretar e discutir, pois seguem a lógica usualmente empregada em classificação sistemática, não apenas de solos, mas de outras áreas do conhecimento, como a biologia (McKenzie & Ryan, 1999; Miller & Franklin, 2001; Xu et al., 2005).

Segundo Moore et al. (1991), árvores de decisão não necessitam de qualquer suposição prévia sobre a distribuição dos dados em cada classe, são invariantes com relação a transformações das variáveis explanatórias, lidam satisfatoriamente com dados perdidos e anômalos, são maleáveis na captura de comportamentos não aditivos e não lineares e podem ser facilmente atualizadas conforme aumente a coleta de dados. McKenzie & Ryan (1999) acrescentam que a capacidade de modelar relações não aditivas e não lineares de uma forma relativamente simples é particularmente importante para dados pedológicos, onde interações entre a variável resposta e as variáveis ambientais explanatórias freqüentemente são condicionadas por outras variáveis explanatórias.

Árvores de decisão também são capazes de processar grandes volumes de dados praticamente sem a necessidade de pré-processamento ou transformação, relacionar variáveis discretas e contínuas e validar os resultados através de métodos estatísticos. Além disso, possibilitam usar dados de diferentes tipos e fontes, como imagens de sensoriamento remoto, MNT e mapas existentes, permitem integrar no mesmo modelo dados mensurados em diferentes escalas e necessitam de tempo de treinamento e de processamento reduzido (Miller & Franklin, 2001; Pal & Mather, 2003).

Elas são tolerantes à sobreposição ou redundância de dados e são capazes de produzir resultados de boa acurácia mesmo com conjuntos reduzidos de amostras de treinamento. Sua capacidade de processar grandes volumes de dados sem interferência humana ainda lhe confere o potencial de recuperar conhecimento pedológico a partir de mapas existentes, através das regras de classificação, e de descobrir estruturas de relacionamentos complexos, tornando-as uma eficiente ferramenta de mineração de dados (Foody, 1995; Gahegan, 2003, Henderson et al., 2005).

As árvores de decisão têm sido empregadas com sucesso para várias finalidades em mapeamento digital de solos, embora ainda menos que outros métodos. Qi & Zhu (2003), por exemplo, utilizaram mineração de dados

por árvores de decisão para extrair o conhecimento contido em mapas de solos existentes. Os autores recuperaram o modelo de relação solo-paisagem destes mapas na forma de regras de classificação e compararam com critérios estabelecidos por pedólogos com experiência de campo local, encontrando alta coincidência. Aplicando as regras recuperadas para predizer novamente as classes de solos, os resultados apresentaram uma concordância média com dados de campo medida com o índice *Kappa* de 0,83. Adicionando mais três variáveis preditoras o índice *Kappa* aumentou para 0,865. Os autores afirmam que o conhecimento sobre as relações solo-paisagem recuperado do mapa existente e reelaborado na forma de regras de classificação poderão ser empregados para treinar novos pedólogos ou para produzir novos mapas de solos onde as mesmas relações sejam válidas.

Zhou et al. (2004) usaram abordagem semelhante para comparar árvores de decisão e um classificador Bayesiano em uma área na província de Zhejiang, na China. Ao reaplicar as regras obtidas do mapa existente os autores obtiveram uma concordância média da predição de 81,3% para a árvore de decisão e de 79,3% para o método de classificação Bayesiano, mas a árvore de decisão obteve concordância alta em todas as classes individuais estimadas enquanto o segundo método subestimou algumas classes.

Hansen et al. (2009) usaram árvores de decisão para classificar unidades de solo-paisagem em uma área de 2.214 km² na região central de Uganda, na África. Os autores utilizaram 66 variáveis explanatórias derivadas de imagens de sensoriamento remoto e de MNT para estimar um mapa com 20 metros de resolução espacial. A avaliação de acurácia foi realizada com base em 190 polígonos homogêneos e 46 transectos observados em campo, encontrando exatidão geral de 75,5% e índice *Kappa* de 0,67. O método foi considerado promissor, especialmente para mapear grandes extensões onde há carência de mapas e informações de solos.

Elnagar & Noller (2010) empregaram árvore de decisão, MNT e imagens de sensoriamento remoto para classificar graus de salinização do solo em uma área de 1.160 km² com diferentes tipos de solos salinos no sudeste do Oregon, Estados Unidos. A avaliação de acurácia foi realizada com base em 210 pontos observados em campo, classificados de acordo com os resultados das análises em laboratório. A comparação do mapa de classes de salinidade

estimado com as classes dos pontos de campo mostrou concordância muito alta, com exatidão geral de 98,4% e índice *Kappa* de 0,9. Após ajustes em parâmetros para melhorar o desempenho da árvore de decisão, a exatidão geral aumentou para 98,8% e o índice *Kappa* para 0,92. Os autores concluem que árvores de decisão podem contribuir muito no mapeamento de solos salinos, tarefa importante para adequar o manejo e monitorar a degradação pelo uso inadequado.

2.3.2. Uso de atributos do terreno

As relações entre os solos e as formas da paisagem têm sido por longo tempo a base para identificar e descrever a ocorrência e distribuição espacial dos solos (Ippolit et al., 2003). O termo paisagem é usualmente empregado para descrever o relevo (ou topografia) através da síntese de algumas características do terreno, como posição na encosta e forma da superfície (Debella-Gilo et al., 2007). O relevo é amplamente reconhecido como um fator pedogenético dominante em nível local, influenciando os solos e suas propriedades. Ele controla a distribuição, redistribuição e acumulação de água e de energia na paisagem, o que, por sua vez, determina em grande parte a textura, a drenagem e o desenvolvimento do perfil (MacMillan et al., 2008).

O relevo desempenha um papel crucial porque, com exceção do tempo, ele modifica o papel que os outros fatores exercem na gênese do solo. Ele modifica a distribuição dos processos hidrológicos e erosionais (isto é, conteúdo de água, escoamento superficial e sedimentação) e a temperatura do solo. O relevo desse modo afeta a distribuição da intemperização de minerais, lixiviação, erosão, sedimentação, decomposição, formação dos horizontes e, por fim, os atributos do solo (Moore et al., 1991; Moore et al., 1993). Ele determina se o solo irá ou não se desenvolver em um determinado espaço e, caso se desenvolva, determina o tipo de solo resultante e, dentro dele, a variação de propriedades individuais (Debella-Gilo et al., 2007).

Assim como nos levantamentos convencionais, grande parte dos estudos em mapeamento digital de solos baseia-se nas relações das classes ou propriedades do solo com as características da paisagem. A diferença é que, enquanto os primeiros usam uma caracterização nominal e qualitativa da topografia, no mapeamento digital as características topográficas são representadas por atributos quantitativos calculados a partir de MNT. Existem

várias estruturas para armazenamento e representação de MNT, sendo mais utilizada a estrutura *raster* (matricial), onde o valor da altitude do terreno em cada local é armazenado na forma de *pixels* (células) quadrados. Esta forma de armazenamento facilita o processamento e a integração com outros dados espaciais porque muitos Sistemas de Informação Geográfica também usam estruturas *raster* (Moore et al., 1993; Burrough & McDonnell, 1998).

Além de fornecer diretamente valores de elevação, a modelagem digital do terreno permite derivar quantitativamente atributos topográficos e hidrológicos do terreno, que podem ser usados não apenas para estimativas de classes ou propriedades de solos, mas também como indicadores de processos ambientais (Moore et al., 1993; McKenzie & Ryan, 1999; Burrough et al., 2000; Wilson & Gallant, 2000; Li et al., 2005; Debella-Gilo et al., 2007; MacMillan et al., 2008).

Os atributos topográficos podem ser divididos em primários e secundários (ou compostos). Os atributos primários são calculados diretamente do MNT e incluem variáveis como a altitude, declividade, orientação, curvatura de perfil e planar, comprimento de fluxo e área de contribuição (ou acúmulo de fluxo). Os atributos secundários são calculados a partir de diferentes combinações dos primários e podem ser usados para caracterizar a variabilidade espacial de processos específicos na paisagem, tanto a partir de relações empíricas quanto pela aplicação de equações que descrevem a física por trás do processo em questão. Nesta categoria encontram-se diversos índices, como o índice de umidade topográfica e o índice de poder de escoamento, entre outros (Moore et al., 1991; Moore et al., 1993).

2.3.2.1. Declividade

A declividade corresponde à inclinação da superfície, o seu gradiente, e é definida como o ângulo entre o plano horizontal e um plano tangente à superfície do terreno e pode ser expressa em graus ou em percentagem. A declividade está relacionada ao regime hídrico do solo, afetando o fluxo de água superficial e subsuperficial, a magnitude e a velocidade do escoamento superficial, a infiltração da água da chuva, o conteúdo de água no solo, a capacidade de uso do terreno, entre outros (Wilson & Gallant, 2000). Considerando-se os demais atributos uniformes, os solos tendem a se desenvolver mais rapidamente e a serem mais profundos

em áreas planas em comparação com áreas íngremes, pois os regimes de umidade são mais favoráveis e os materiais tendem a acumular mais em áreas planas (Debella-Gilo et al., 2007).

2.3.2.2. Orientação

A orientação é a direção para a qual a normal à superfície do terreno está voltada. Usualmente é expressa na forma do azimute, em graus decimais, com valores que variam de 0° a 360° no sentido horário a partir do Norte (Wilson & Gallant, 2000). A orientação da superfície modifica a influência da encosta, expondo ou ocultando a superfície e afetando a interceptação de radiação solar. Conseqüentemente, modifica a temperatura da superfície e os regimes de umidade, agindo na evapotranspiração e na direção dos fluxos de água laterais, entre outros (Romano & Chirico, 2004).

2.3.2.3. Curvatura

A curvatura fornece um valor relacionado à concavidade ou convexidade da superfície na direção considerada. Segundo Wilson & Gallant (2000), é possível derivar várias medidas de curvatura de um MNT, sendo mais comuns a curvatura de perfil, também denominada de vertical ou máxima, e a curvatura planar, também denominada de horizontal, tangencial ou mínima. A curvatura de perfil expressa a curvatura da superfície do terreno na direção da sua orientação e a curvatura planar expressa a curvatura ortogonalmente à direção da orientação. A curvatura da superfície está relacionada à velocidade do escoamento superficial (aceleração ou desaceleração), à velocidade e direção dos fluxos laterais e às taxas de erosão e deposição, entre outros (Romano & Chirico, 2004). A curvatura é tão importante quanto a declividade, porque a concavidade e a convexidade da superfície inclinada regulam o armazenamento e o fluxo de água e de materiais sólidos ao longo da encosta (Debella-Gilo et al., 2007).

2.3.2.4. Direção de fluxo

A direção de fluxo indica a direção para a qual segue o fluxo que deixa uma determinada célula do MNT. O princípio fundamental na determinação da direção de fluxo é que a água escoar de uma parte mais alta para uma parte mais baixa no terreno (Li et al., 2005). Segundo Wilson & Gallant (2000), existem vários algoritmos de direção de fluxo, sendo que os

mais simples adotam a orientação da superfície e os mais sofisticados podem considerar múltiplos caminhos de fluxo. A Figura 6 mostra um esquema da determinação da direção do fluxo para um conjunto de células de um MNT (a), onde os valores representam a elevação do terreno em metros, e a representação da direção do fluxo sobre uma área maior (b).

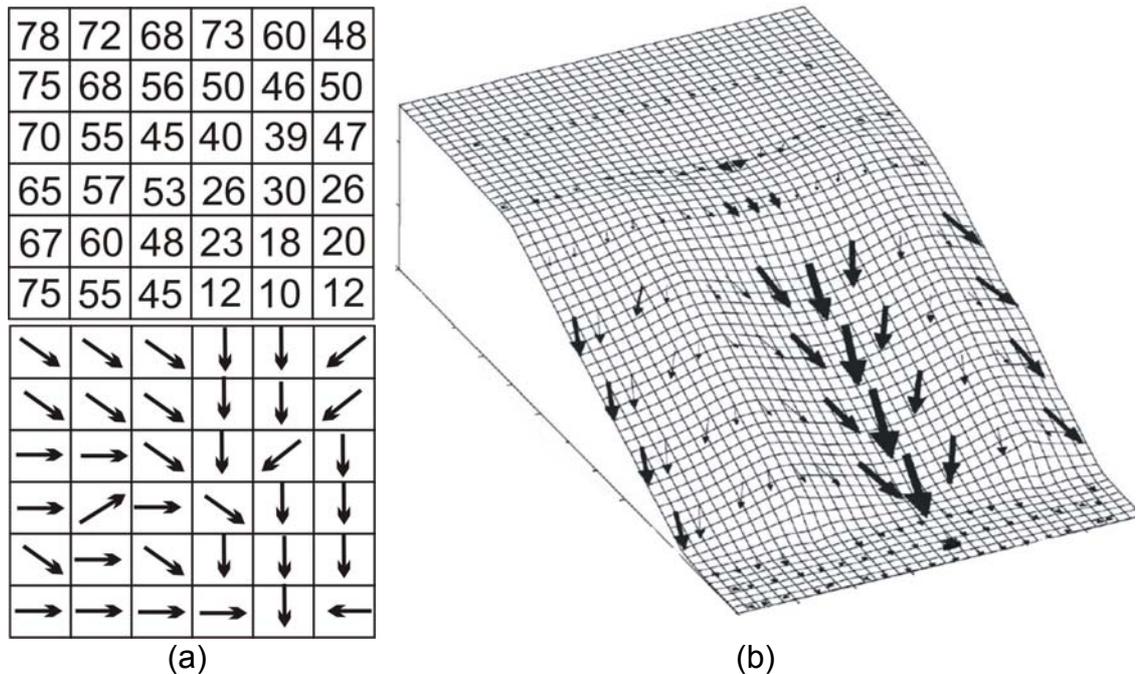


Figura 6. (a) Esquema da determinação da direção do fluxo sobre um MNT (Li et al., 2005); (b) Representação da direção de fluxo sobre uma área maior (Pennock, 2003).

2.3.2.5. Acúmulo de fluxo

Dispondo-se da direção de fluxo é possível estimar o acúmulo de fluxo em cada célula. O acúmulo de fluxo, ou fluxo acumulado, é expresso como o número de células a montante que contribuem com o fluxo para uma determinada célula (Dobos et al., 2006) e está relacionado ao regime de umidade do solo, o tipo de escoamento, laminar ou canalizado, entre outros fatores. Representa o tamanho da área de contribuição para uma determinada célula e constitui uma medida crítica no cômputo de vários atributos hidrológicos e índices de solo-paisagem.

2.3.3. Avaliação da acurácia de mapas

Independente das técnicas e metodologias de mapeamento digital e do tipo de resultado produzido, mensurar de alguma maneira a acurácia dos mapas gerados tem se tornado uma preocupação crescente. Conhecer a

qualidade é fundamental para orientar a tomada de decisões sobre o uso do mapa, pois permite avaliar sua adequação às demandas de diferentes usuários (Rossiter, 2004b; Carré et al., 2007).

Segundo Foody (2002), citando Congalton (1991), a avaliação de acurácia evoluiu consideravelmente desde os primeiros estudos de mapeamento, passando por vários estágios com sucessivo incremento no detalhe e no rigor das análises. No primeiro estágio a avaliação implicava em um alto grau de subjetividade, pois se baseava em uma inspeção visual do mapa e eventualmente a comparação qualitativa com alguma informação de referência. Caso sua aparência parecesse boa ou correta ao avaliador, ele era considerado adequado.

O segundo estágio caracterizou-se pela busca de medidas mais objetivas, utilizando como indicador a relação entre a extensão em superfície das classes no mapa e a respectiva extensão no terreno ou em outros dados de referência. A maior limitação dessa abordagem consiste no fato de que freqüentemente o mapa estimado poderia reproduzir a extensão correta das classes, mesmo a sua localização estando incorreta. Tanto na primeira quanto na segunda abordagem, a qualidade aparente do mapa apresenta o risco de ocultar a sua qualidade real.

O terceiro estágio também envolveu a comparação das classes identificadas no mapa e no terreno, mas considerando um conjunto específico de locais para comparação. Assim, a localização das classes passou a fazer parte da avaliação de acurácia na forma da porcentagem de casos corretamente classificados, medida também conhecida como acurácia geral ou exatidão geral.

O quarto estágio é um refinamento do terceiro e se caracteriza por um uso mais aprofundado da correspondência entre as classes mapeadas e as classes de referência, tendo como base a matriz de erro. A partir da matriz de erro, além da exatidão geral, podem ser computadas medidas de acurácia mais consistentes, como o índice *Kappa* de concordância (Cohen, 1960), adotado para avaliação de mapas em diversas áreas de aplicação.

2.3.3.1. Matriz de erro

A matriz de erro (Tabela 1) relaciona a quantidade de unidades amostrais associadas a uma dada categoria durante o processo de

classificação e a categoria correta à qual pertencem essas unidades (Congalton, 1991; Foody, 2002). As colunas representam os dados de referência e as linhas representam o resultado da predição. O total da coluna quantifica a extensão da respectiva classe nos dados de referência e o total da linha quantifica a extensão da respectiva classe no mapa estimado. A diagonal principal representa os casos onde a classe estimada e a classe de referência são concordantes, enquanto as demais células representam os dois tipos de erros associados ao sistema de classificação: os erros de inclusão e os erros de omissão.

Os erros de inclusão são calculados ao longo das linhas e quantificam os casos nos quais outras classes foram atribuídas a uma classe particular na predição. São obtidos dividindo-se a soma dos casos fora da diagonal principal na linha pelo total da linha. Os erros de omissão são calculados ao longo das colunas e quantificam os casos pertencentes a uma determinada classe no mapa original que foram atribuídos a outras classes na predição. São obtidos dividindo-se a soma dos casos fora da diagonal principal na coluna pelo total da coluna (Campbell & Mortenson, 1989; Rossiter, 2001; Foody, 2002).

Tabela 1. Esquema de uma matriz de erros para caso com três categorias (A, B e C). Σ_e = Total estimado, EI = erros de inclusão, AU = acurácia do usuário, Σ_r = total da referência, EO = erros de omissão, AM = acurácia do mapeador.

	A	B	C	Σ_e	EI (%)	AU (%)
A	X_{11}	X_{12}	X_{13}	$\Sigma_{X_{1i}}$ ($i=1,2,3$)	$\frac{(X_{12}+X_{13}) * 100}{\Sigma_{X_{1i}}}$	$\frac{X_{11} * 100}{\Sigma_{X_{1i}}}$
B	X_{21}	X_{22}	X_{23}	$\Sigma_{X_{2i}}$ ($i=1,2,3$)	$\frac{(X_{21}+X_{23}) * 100}{\Sigma_{X_{2i}}}$	$\frac{X_{22} * 100}{\Sigma_{X_{2i}}}$
C	X_{31}	X_{32}	X_{33}	$\Sigma_{X_{3i}}$ ($i=1,2,3$)	$\frac{(X_{31}+X_{32}) * 100}{\Sigma_{X_{3i}}}$	$\frac{X_{33} * 100}{\Sigma_{X_{3i}}}$
Σ_r	$\Sigma_{X_{j1}}$ ($j=1,2,3$)	$\Sigma_{X_{j2}}$ ($j=1,2,3$)	$\Sigma_{X_{j3}}$ ($j=1,2,3$)	$\Sigma_{X_{ij}}$ ($i,j=1,2,3$)		
EO (%)	$\frac{(X_{21}+X_{31}) * 100}{\Sigma_{X_{j1}}}$	$\frac{(X_{12}+X_{32}) * 100}{\Sigma_{X_{j2}}}$	$\frac{(X_{13}+X_{33}) * 100}{\Sigma_{X_{j3}}}$			
AM (%)	$\frac{X_{11} * 100}{\Sigma_{X_{j1}}}$	$\frac{X_{22} * 100}{\Sigma_{X_{j2}}}$	$\frac{X_{33} * 100}{\Sigma_{X_{j3}}}$			

Originalmente as matrizes de erro foram concebidas para identificar e quantificar os diferentes tipos de erros que, uma vez conhecidos, podem ser tratados ou eliminados a fim de melhorar a qualidade do mapa (Foody, 2002). Entretanto, freqüentemente os usuários necessitam saber a confiabilidade de uma classe particular para avaliar se ela pode ser usada para o propósito desejado. Por essa razão, matrizes de erro muitas vezes são usadas para expressar a acurácia das classes em vez dos respectivos erros. A acurácia do usuário é calculada ao longo da linha dividindo-se o número de casos corretamente classificados pelo total da linha. A acurácia do mapeador é calculada ao longo da coluna dividindo-se o número de casos corretamente classificados pelo total da coluna (Filippi & Jensen 2006; Elnaggar, 2007).

Outro aprimoramento sugerido por Rossiter (2001) é usar matrizes de erro ponderadas. Nesse caso, assume-se que as classes não possuem todas a mesma importância para o usuário, porque algumas podem ser mais relevantes para o propósito em questão. Dessa forma, erros elevados (ou baixa acurácia) em uma classe de pouca importância são menos relevantes que erros menores (ou acurácia maior) em classes de maior importância. Entretanto, como a atribuição de pesos para as classes é um processo subjetivo convém que o próprio usuário final execute essa tarefa.

2.3.3.2. Exatidão geral

A partir da matriz de erro, que expressa os erros por classe é possível derivar várias medidas de acurácia relativas ao mapa estimado como um todo. A mais simples é a exatidão geral, expressa como a percentagem de *pixels* corretamente classificados na área mapeada, ou seja, a proporção entre o número de *pixels* corretamente classificados e o número total de *pixels* do mapa (Foody, 2002). É calculada dividindo-se a soma dos elementos da diagonal principal pelo total de elementos da matriz, conforme a Equação 5 abaixo:

$$E_g = \frac{X_{11} + X_{22} + X_{33}}{\sum X_{ij} (i,j=1,2,3...)} * 100 \quad (5)$$

Onde E_g é a exatidão geral, X_{11} , X_{22} e X_{33} são os elementos da diagonal principal, correspondentes aos casos corretamente classificados, e $\sum X_{ij}$ é a soma de todos os elementos da matriz.

2.3.3.3. Índice *Kappa*

Embora sejam de indiscutível valor informativo, matrizes de erro e medidas simples como os erros de inclusão e de omissão e a exatidão geral, não levam em conta os casos em que as concordâncias entre o mapa estimado e o mapa convencional de solos usado como referência podem ocorrer puramente devido ao acaso (Congalton, 1991; Pontius, 2000). Para compensar as concordâncias devido ao acaso a estatística mais freqüentemente empregada é o índice *Kappa* (Cohen, 1960), uma técnica multivariada discreta usada para mensurar a concordância entre categorias estimadas e categorias de referência. O cálculo do índice *Kappa* leva em conta o número total de acertos, os erros de omissão e os erros de inclusão e é efetuado conforme a Equação 6 abaixo (Elnaggar, 2007):

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} * x_{+i})} \quad (6)$$

onde *K* é o índice *Kappa*, *N* é o número total de posições na matriz de erro, *r* é o número de linhas na matriz, *x_{ii}* são as entradas diagonais da matriz, *x_{i+}* e *x_{+i}* indicam a soma da linha *i* e a soma da coluna *i* da matriz, respectivamente.

Dessa forma, o índice *Kappa* expressa o grau de concordância entre os dados estimados e os dados de referência, após a concordância devido ao acaso ter sido eliminada. Em avaliações de mapeamentos, o índice *Kappa* geralmente varia de 0 a 1, sendo que zero (0) indica ausência de concordância, ou seja, a estimativa e a referência são estatisticamente independentes (Elnaggar, 2007), e um (1) indica total concordância. Valores negativos de *Kappa* podem ocorrer quando a concordância encontrada é menor do que aquela devido ao acaso, embora isso raramente ocorra em mapeamentos de solos (Congalton, 1991; Pontius, 2000).

Em avaliações de mapeamentos, não apenas de solos, valores de índice *Kappa* superiores a 0,5 são considerados satisfatórios, mas Luoto & Hjort (2005) sugerem o uso de uma escala para enquadrar a concordância. Proposta originalmente por Landis & Koch (1977), essa escala adota os

seguintes intervalos: 0,81 – 1,00 = concordância muito forte; 0,61 – 0,80 = concordância forte; 0,41 – 0,60 = concordância moderada; 0,21 – 0,40 = concordância fraca; 0,00 – 0,20 = sem concordância.

2.3.4. Questões relacionadas à avaliação de acurácia

Atualmente a matriz de erro, a exatidão geral e o índice *Kappa* constituem o núcleo da avaliação de acurácia. Os três indicadores são igualmente importantes, pois fornecem informações complementares sobre a qualidade do mapa avaliado. A matriz de erro possibilita identificar as classes mais concordantes e aquelas em que a predição está errando mais, enquanto a exatidão geral e o índice *Kappa* fornecem medidas sobre a concordância do mapa estimado como um todo (Rossiter, 2001).

A possibilidade de avaliar a acurácia dos mapas digitais de solos simultaneamente à sua geração representa uma vantagem em relação aos levantamentos convencionais. Nestes, raramente medidas quantitativas de acurácia estão disponíveis e, caso sejam feitas, possivelmente graus de acurácia abaixo do esperado sejam encontrados em muitos casos (Liu & Zhu, 2009; Lopez, 2009). Segundo MacMillan (2008), os mapas estimados através de técnicas de mapeamento digital tem apresentado acurácia igual ou superior aos mapas convencionais de solos preparados para a mesma área.

Zhu (2000), por exemplo, empregando redes neurais na predição de classes de solos na escala 1:10.000, obteve exatidão geral de 77% ao comparar o mapa estimado com 64 pontos observados em campo, enquanto o mapa convencional disponível para a mesma área apresentou exatidão geral de 61%. Shi et al. (2004), usando lógica de conjuntos difusos para estimar classes de solos em uma bacia hidrográfica de 5 km², obtiveram exatidão geral de 86% na identificação das classes pelo mapa estimado contra 46% do mapa convencional, usando dados de referência observados em 57 pontos.

Chagas et al. (2007), empregando uma rede neural MLP e o algoritmo da Máxima Verossimilhança Gaussiana (MAXVER) para estimar classes de solos na escala 1:50.000 em uma região montanhosa do Rio de Janeiro, obtiveram exatidão geral e índice *Kappa* respectivamente de 70,83% e 0,678 para a rede neural MLP e de 50,69% e 0,46 para o algoritmo MAXVER. O mapa convencional de solos existente para a área de estudo apresentou exatidão geral de 52,77% e índice *Kappa* de 0,469. Ao comparar os mapas

estimados com o mapa convencional existente obtiveram uma concordância de apenas 37,15% para a rede neural MLP e de 22,3% para o algoritmo MAXVER.

Embora sejam de uso corrente e tenham grande importância como indicadores de qualidade, a matriz de erro, a exatidão geral e o índice *Kappa* não caracterizam completamente os diversos tipos de erros envolvidos no processo de mapeamento (Rossiter, 2001). Ainda há muitas incertezas associadas com o significado e a interpretação dessas medidas de qualidade, o que pode levar a uma comunicação imperfeita da qualidade dos mapas para o usuário (Johnston & Timlin, 2000; Foody, 2002). Rossiter (2001) cita a necessidade de detalhar mais a caracterização dos diversos tipos de erros envolvidos no processo de mapeamento, que não são abrangidos pelas medidas usuais. Segundo o autor, matrizes de erro podem fornecer muito mais informações do que as correntemente utilizadas.

Foody (2002) levanta dúvidas sobre a adequação do índice *Kappa*, uma vez que ao tentar eliminar o efeito do acaso ele subestima a qualidade real dos mapas. Há também questões relacionadas à amostragem de pontos para a avaliação de acurácia, tanto com relação ao método de seleção de pontos quanto à densidade de amostragem e sua relação com a escala de mapeamento, à acurácia dos dados de referência, tipos, magnitude e distribuição espacial dos erros, entre outras. Essas questões demonstram que há a necessidade de mais pesquisas para planejamento da amostragem, estimativa de qualidade dos mapas e treinamento de pessoal para trabalhar com esta nova forma de mapeamento do solo (McBratney et al., 2003).

3. MATERIAL E MÉTODOS

3.1. Caracterização da área de estudo

3.1.1. Localização

A área de estudo situa-se no Vale dos Vinhedos, na região de produção vitivinícola denominada de Serra Gaúcha, localizada na porção nordeste do Estado do Rio Grande do Sul, aproximadamente entre as latitudes 28°30'S e 29°30'S e longitudes 50°40'W e 52°05'W (Figura 7). Segundo Hoffmann et al. (2005), a Serra Gaúcha engloba 31 municípios, distribuídos em três regiões fisiográficas, com a maior parte dos municípios pertencendo à Encosta Superior do Nordeste (Fortes, 1979). De maneira geral, as condições naturais da região são heterogêneas, com relevo complexo (grandes variações de altitude, declividade e exposição solar) e grande variabilidade de tipos de solos (Sarmiento et al., 2006).

A Serra Gaúcha é a principal região produtora de vinhos finos do Rio Grande do Sul, Estado que responde por 95% da produção nacional de vinhos finos, com uma área de cerca de 5.000 hectares plantados com uvas viníferas. Com origem vinculada à imigração italiana, a viticultura da Serra Gaúcha é uma atividade de pequenas propriedades, com média de 15 ha de área total e 2,5 ha de vinhedos, topografia acidentada, pouca mecanização e predomínio da mão-de-obra familiar (Protas et al., 2004).

O Vale dos Vinhedos possui uma superfície de aproximadamente 8.279 hectares e divide-se entre os municípios de Bento Gonçalves (62,4%), Garibaldi (30,7%) e Monte Belo do Sul (6,9%). Fruto de uma estratégia voltada à tipificação e valorização da produção vitivinícola da Serra Gaúcha, o Vale dos Vinhedos foi a primeira indicação geográfica do Brasil, implantada inicialmente em nível de Indicação de Procedência (Tonietto & Mandelli, 2005). Atualmente encontra-se em andamento o aprimoramento para Denominação de Origem (DO), nível que exige um maior detalhamento dos fatores que afetam a produção vitivinícola, entre os quais o solo.

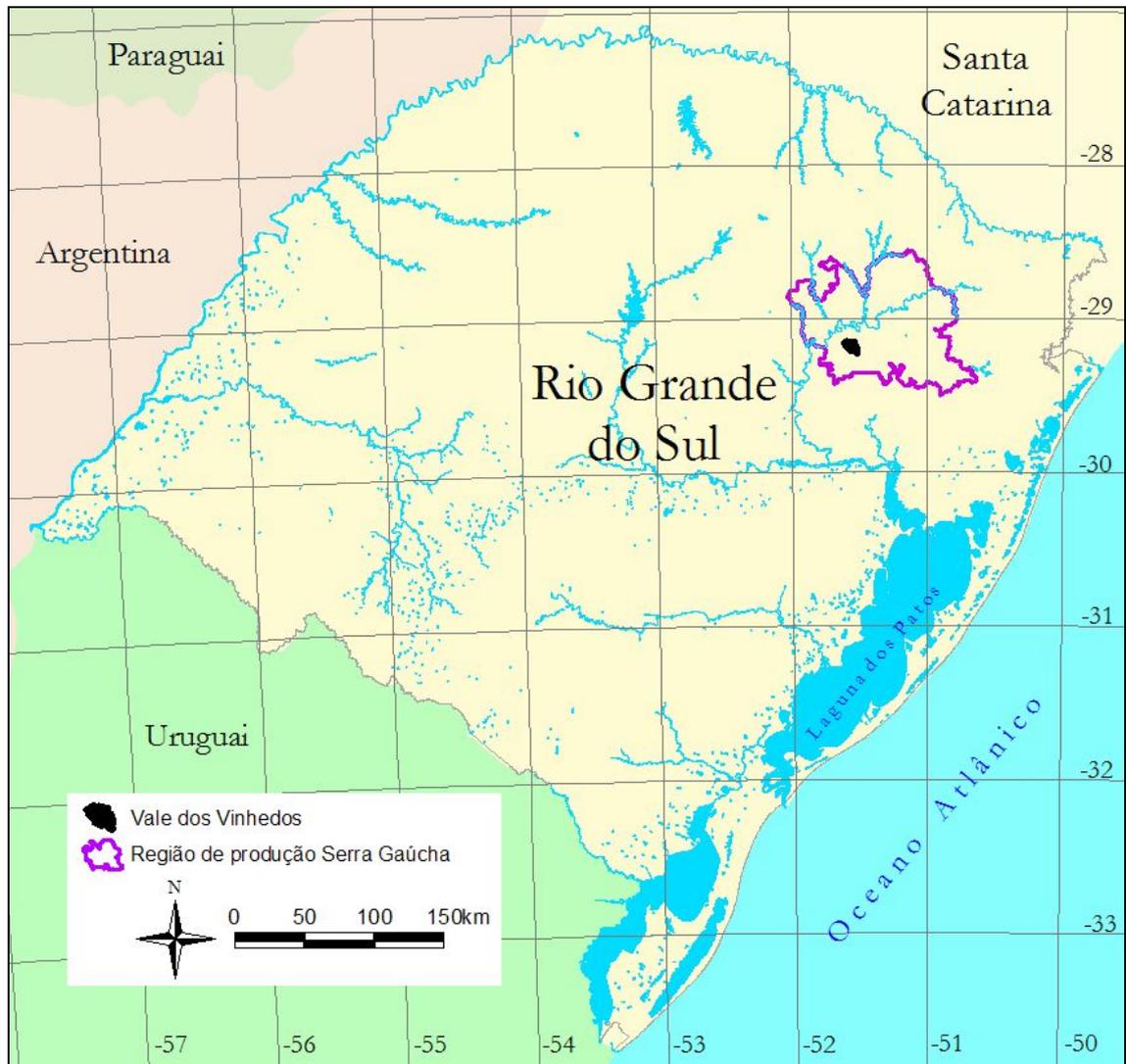


Figura 7. Localização da região de produção vitivinícola da Serra Gaúcha e do Vale dos Vinhedos no Rio Grande do Sul.

A área de estudo compreende uma superfície de 673,5 ha, correspondente ao retângulo de uma carta do levantamento detalhado de solos (escala 1:5.000) do Vale dos Vinhedos (Sarmiento et al., 2008). Os limites deste retângulo situam-se aproximadamente entre as coordenadas UTM (fuso 22) 444.050 m E, 446.540 m E, 6.772.460 m N e 6.775.290 m N. A área encontra-se situada integralmente dentro dos limites da Indicação de Procedência Vale dos Vinhedos e do município de Bento Gonçalves (Figura 8).

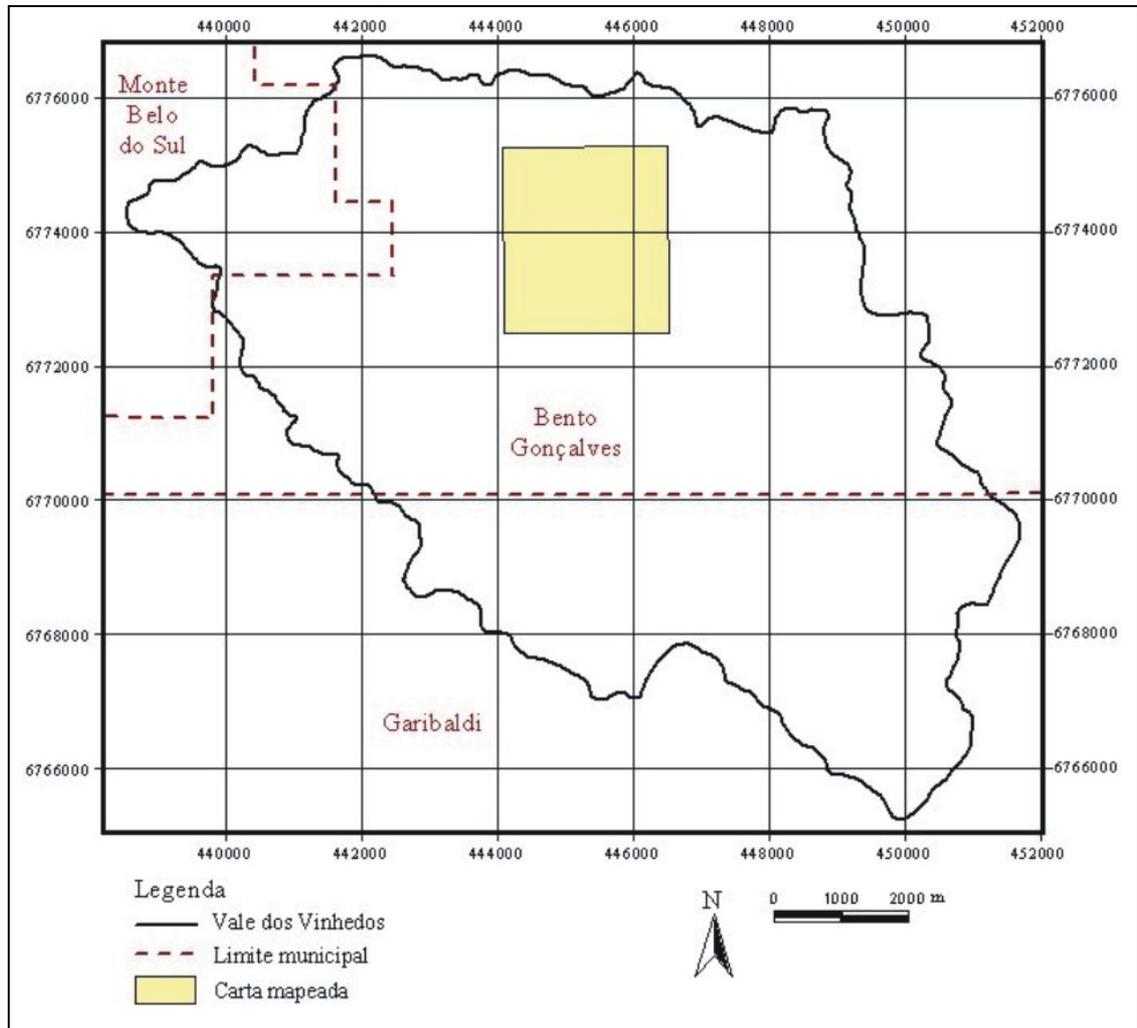


Figura 8. Localização da área de estudo.

3.1.2. Clima

A Serra Gaúcha, como o restante do estado do Rio Grande do Sul, encontra-se na primavera-verão sob domínio da Massa Tropical Marítima, responsável por chuvas intensas e de curta duração. No outono-inverno, ocorre a entrada de ciclones migratórios polares associados à Massa Polar Marítima, trazendo chuvas menos intensas e de duração mais prolongada. Por essa razão, o clima é úmido e, devido à altitude, de temperaturas amenas (Hasenack & Ferraro, 1989).

O clima da Serra Gaúcha caracteriza-se pela homogeneidade pluviométrica e unidade no domínio “do clima mesotérmico do tipo temperado” (Nimer, 1989), devido às condições do relevo, da localização e da dinâmica da

circulação atmosférica. A temperatura apresenta comportamento uniforme, mas não homogêneo, pois é condicionada, sobretudo, pelo relevo. Na maior parte dos patamares do planalto as médias mensais situam-se entre 16°C e 18°C, mas nas áreas mais elevadas das escarpas do Planalto das Araucárias podem ser inferiores a 10°C nos meses de inverno acentuado ou rigoroso (Nimer, 1989). As geadas são freqüentes no inverno, podendo ocorrer também na primavera e no outono. A penetração da massa de ar polar atlântica, muito fria e úmida, pode provocar, ao elevar-se sobre as áreas de maior altitude, a precipitação de neve.

De acordo com as normais climatológicas da estação de Bento Gonçalves (Embrapa, 2009), situada próxima à área de estudo, a temperatura média anual é de 17,2°C, sendo janeiro o mês mais quente, com temperatura média de 21,8°C, e junho o mais frio, com temperatura média de 12,8°C. A precipitação total anual é de 1.736 mm, sendo maio o mês mais seco, com precipitação acumulada de 107 mm, e setembro o mais chuvoso, com precipitação acumulada de 185 mm (Figura 9). Segundo a classificação climática de Köppen, a área de estudo apresenta um clima do tipo Cfb (Moreno, 1961), subtropical com verão ameno, com temperatura média do mês mais frio entre -3°C e 18°C, temperatura média do mês mais quente inferior a 22°C e precipitação pluvial bem distribuída ao longo do ano.

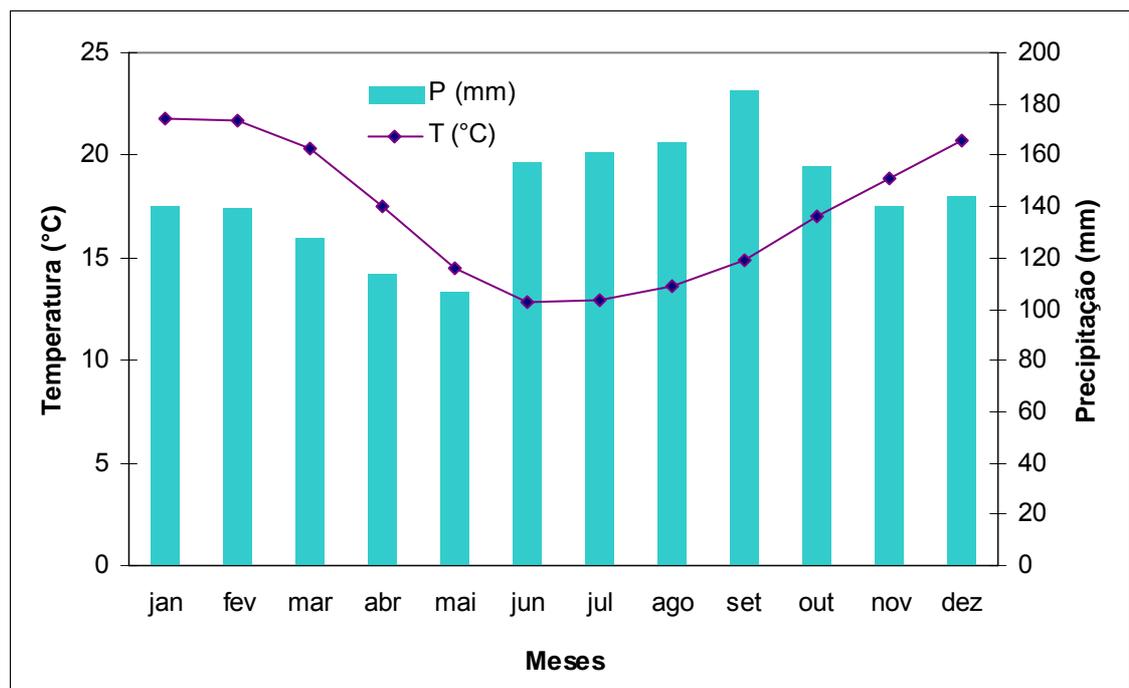


Figura 9. Temperatura média mensal e precipitação pluvial total mensal em Bento Gonçalves, RS (1961 – 1990).

3.1.3. Geologia

A origem geológica da região da Serra Gaúcha integra a Formação Serra Geral da Série São Bento, na Bacia Rio do Paraná (IBGE, 1986; IBGE, 1990), cronologicamente pertencente aos períodos triássico e jurássico, estendendo-se até o cretáceo inferior, da era mesozóica.

A Formação Serra Geral constitui-se de uma sucessão de derrames de rochas efusivas, de composição predominantemente básica, compreendendo derrames de basalto, andesito, além de brechas vulcânicas e sedimentares, diques e soleiras de diabásio e corpos de arenitos interderrames (IBGE, 1986). Mais de 60% do Estado do Rio Grande do Sul foi coberto por esses derrames, que atingiram ainda parte dos Estados de Santa Catarina, Paraná, São Paulo, Mato Grosso do Sul e Mato Grosso, além do norte do Uruguai e o leste da Argentina e do Paraguai. No Rio Grande do Sul o derrame ocorre desde cotas inferiores a 100 metros até aproximadamente 1.200 metros de altitude (Flores, 2009).

A seqüência inferior da Formação Serra Geral é constituída predominantemente de rochas efusivas básicas, comumente basaltos e fenobasaltos, com diques e corpos tabulares de diabásio associados (IBGE, 1986). Os basaltos apresentam uma variedade de cores que grada de cinza-escuro ao negro, com tonalidades esverdeadas. As colorações escuras devem-se principalmente à granulação fina e à abundante presença de minerais ferromagnesianos opacos e de vidros (Falcade & Mandelli et al., 1999). Ocasionalmente, entre as lavas, ocorrem lentes e camadas de arenitos interderrames de origem eólica, de coloração rósea, com estratificação cruzada tangencial e brechas constituídas por fragmentos de basalto e arenitos cimentados por lava basáltica. A espessura dos arenitos interderrames pode variar de alguns centímetros até dezenas de metros (IBGE, 1986).

Normalmente capeando as efusivas básicas, na seqüência superior dos derrames, ocorre um domínio relativo de rochas efusivas de composição ácida, constituídas por riolitos, dacitos e basaltos pórfiros, entre outros (IBGE, 1986). As rochas riolíticas, quando alteradas, exibem coloração com tons de cinza-claro e amarelados, tornando-se até avermelhados quando impregnadas por óxidos de ferro. Ocorrem principalmente nas partes mais elevadas, especialmente na região dos Campos de Cima da Serra. Os basaltos pórfiros

apresentam coloração variando de cinza-claro a cinza-escuro (Flores et al., 1999).

Com menor expressão, verifica-se também pequenas ocorrências de sedimentos fluviais e paludais do Quaternário, principalmente nas várzeas dos rios e pequenos corpos d'água (Flores, 2009).

3.1.4. Relevo

A região da Serra Gaúcha situa-se na unidade geomorfológica da Serra Geral, pertencente à região geomorfológica do Planalto das Araucárias, que por sua vez constitui uma subdivisão do domínio morfoestrutural das bacias e coberturas sedimentares, a qual corresponde de modo geral à província geológica do Paraná (IBGE, 1986).

A topografia é ondulada no topo e apresenta bordas escarpadas e recortadas, resultado do processo de dissecação verificado no Sul do Brasil, a partir de alterações climáticas, quando se desenvolveu uma rede de drenagem com maior capacidade de erosão vertical, o que preservou áreas mais elevadas, testemunhos de uma superfície anterior. Apresenta relevos conservados e dissecados de planaltos e suas escarpas, vigorosamente recortadas e festonadas pelo entalhe da drenagem, que secciona as várias seqüências de derrames, deixando nas vertentes abruptas um sucessivo escalonamento de patamares estruturais (Figura 10), modelado conhecido regionalmente por "Região Serrana" (IBGE, 1986; Hoffmann et al., 2005).

Essas superfícies mais elevadas servem de divisores de águas dos inúmeros pequenos córregos que correm para os tributários do Rio Caí, na direção sul e sudeste, e para os tributários do Rio das Antas, na direção norte e noroeste (Hoffmann et al., 2005).



Figura 10. Vista geral do relevo da região.

A área de estudo apresenta um relevo bastante movimentado, variando predominantemente de ondulado a montanhoso, com altitudes que vão de cerca de 260 metros até aproximadamente 630 metros acima do nível do mar. A Tabela 2 mostra a superfície ocupada pelas diferentes fases de relevo, considerando-se os intervalos usualmente utilizados nos levantamentos de solos e modificados por Flores et al. (2006b), quais sejam: plano (0-3%), suave ondulado (3-8%), moderadamente ondulado (8-13%), ondulado (13-20%), forte ondulado (20-45%), montanhoso (45-75%) e escarpado (>75%).

Tabela 2. Superfície ocupada por diferentes fases de relevo na área de estudo.

Classe de declividade	Área (ha)	Proporção (%)
0 a 3%	2,9	0,4
3 a 8%	52,2	7,7
8 a 13%	72,7	10,8
13 a 20%	96,6	14,3
20 a 45%	256,6	38,1
45 a 75%	152,1	22,6
> 75%	40,5	6,0
Total	673,5	100,0

A rede de drenagem apresenta um padrão dendrítico de alta densidade, formada por um grande número de pequenos córregos que fluem em várias direções. Os dois principais cursos d'água que cortam a área de estudo são o arroio Vale dos Vinhedos, em uma pequena extensão na porção Sudoeste, e o arroio Leopoldina, ao longo de toda a extensão na porção central. O arroio Leopoldina forma um vale profundamente encaixado onde ocorrem as declividades mais acentuadas, constatando-se em alguns pontos paredões praticamente verticais (Figura 11).

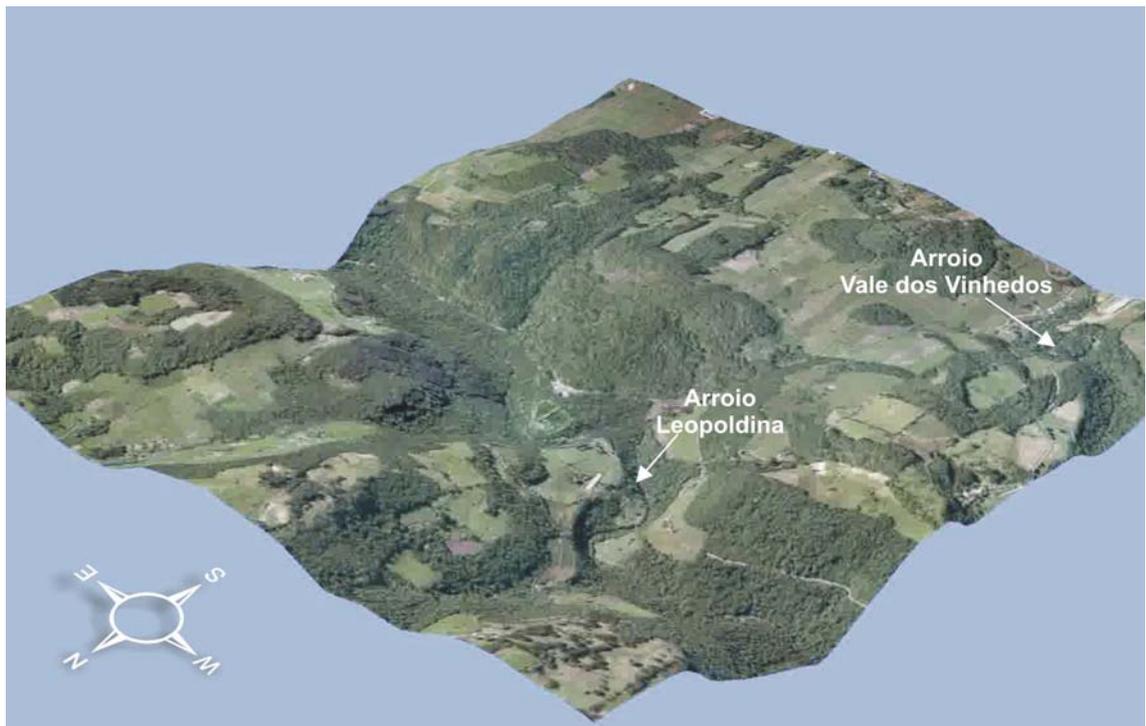


Figura 11. Vista em perspectiva tridimensional da área de estudo, de noroeste para sudeste.

3.1.5. Solos

A região de produção vitivinícola da Serra Gaúcha abrange parcialmente as regiões fisiográficas da Encosta Inferior do Nordeste, da Encosta Superior do Nordeste e dos Campos de Cima da Serra. Em função do relevo movimentado, a distribuição dos solos na região forma um complexo com a predominância de solos rasos e pedregosos (Flores et al., 1999). Entretanto, apesar da limitação imposta pelo relevo e profundidade, as terras são intensamente utilizadas com culturas diversificadas, dada a boa fertilidade natural dos solos e a estrutura de posse da terra, baseada em pequenas propriedades rurais (IBGE, 1986).

Na porção correspondente à Encosta Inferior do Nordeste podem ocorrer Argissolos Vermelhos Distróficos típicos arênicos, Argissolos Vermelho-Amarelos Distróficos abruptos, Argissolos Bruno-Acinzentados Alumínicos típicos, e Chernossolos Argilúvicos Férricos típicos associados a Neossolos Regolíticos ou Litólicos Eutróficos. Entretanto, uma elevada proporção desses Chernossolos foi degradada pelo uso agrícola, resultando na erosão do horizonte A chernozêmico original, transformando-os em Cambissolos Háplicos ou em Luvisolos Háplicos. Nas várzeas encaixadas dos rios Taquari e Caí e de seus afluentes ocorrem Chernossolos Háplicos Órticos típicos (Streck et al., 2008).

Na porção correspondente à Encosta Superior do Nordeste, nos relevos mais acidentados, ocorrem Neossolos Litólicos e Neossolos Regolíticos Distróficos, Neossolos Litólicos Distro-úmbricos fragmentários típicos e Neossolos Regolíticos Distro-úmbricos lépticos ou típicos. Em relevo ondulado, estendendo-se também sobre a porção correspondente aos Campos de Cima da Serra, são encontrados Cambissolos Húmicos Alumínicos típicos associados com Argissolos Bruno-Acinzentados Alíticos abruptos e Neossolos Litólicos, Cambissolos Háplicos Eutróficos e Distróficos, Latossolos Brunos e Nitossolos Brunos (IBGE, 1986; Streck et al., 2008).

Em levantamento semidetalhado na escala 1:50.000 correspondente à Folha SH.22-V-D-II-2 (Bento Gonçalves), Flores et al. (2006a) identificaram a ocorrência de 27 unidades de mapeamento, a maior parte unidades compostas por associações. Associações de Neossolos Litólicos com Cambissolos Háplicos ocupam cerca de 39,5% da área. Cambissolos Háplicos e associações de Cambissolos Háplicos com Neossolos Litólicos e Nitossolos Brunos totalizaram cerca de 24,4% do total. Associações de Argissolos Acinzentados com Cambissolos Háplicos e Neossolos Litólicos e de Argissolos Vermelhos com Chernossolos Háplicos e Neossolos Litólicos somaram 19,4% da superfície. Chernossolos Argilúvicos e associações de Chernossolos Háplicos com Chernossolos Argilúvicos e Neossolos Litólicos ocupam 15,2% e Nitossolos Brunos e associações de Nitossolos Brunos e de Nitossolos Vermelhos com Cambissolos Háplicos somaram cerca de 1,27% da superfície mapeada.

3.1.6. Vegetação

A vegetação original da Serra Gaúcha é resultado não somente das condições atuais de ambiente, mas também das condições reinantes nos períodos glaciários e interglaciários do quartenário (Ab'Saber, 1957). A região caracterizava-se, originalmente, pela Floresta Estacional Decidual nas áreas escarpadas do planalto e pela Floresta Ombrófila Mista, com bosques de *Araucária angustifolia* intercalados por campos, nos topos mais suaves (IBGE, 1986). Nas áreas de contato observa-se uma interpenetração das espécies entre as duas formações vegetais.

A vegetação original da área de estudo corresponde à Floresta Estacional Decidual, formação vegetal onde a queda foliar faz parte do processo de dormência, com uma estrutura organizacional complexa, devido à abundante luminosidade, mas com poucas espécies exclusivas. Os estratos apresentam cobertura bastante densa, os superiores formados por grápia (*Apuleia leiocarpa*), angico-vermelho (*Parapiptadenia rígida*), cabreúva (*Myrocarpus frondosus*), canafístula (*Peltophorum dubium*), pau-marfim (*Balfourodendrom riedelianum*), canelas (*Cryptocarya aschersoniana*, *Ocotea pulchella* e *Ocotea puberula*), entre outras, e os inferiores, além de indivíduos jovens dessas espécies, pelo cincho (*Sorocea bomplandii*), o cotiguá (*Trichilia claussenii*) e gramíneas (IBGE, 1986).

Essa vegetação foi profundamente alterada pela intensa ação antrópica desenvolvida a partir de 1.874, quando iniciou a colonização italiana. Os imigrantes italianos, assentados em pequenas propriedades rurais, promoveram o desmatamento da floresta e a implantação de policulturas, com destaque para o trigo, o milho, o feijão, a videira e ainda pequenas áreas de pastagem, empregando um manejo rudimentar (Hoffman et al., 2005). Contudo, ainda restam algumas áreas de florestas nativas localizadas nas superfícies mais acidentadas e isoladas das escarpas do planalto (Figura 12), embora poucas possam ser consideradas florestas primárias em virtude de já ter havido algum tipo de exploração no passado.



Figura 12. Padrão de uso do solo da região, com exploração agrícola nas áreas menos íngremes (várzeas e patamares dos derrames) e remanescentes de florestas nativas nas áreas mais acidentadas.

3.2. Material utilizado

3.2.1. Mapa convencional de solos

O mapa de solos disponível para a área de estudo foi produzido através de um levantamento convencional de solos, em escala de detalhe, com base em fotografias aéreas e curvas de nível de um aerolevanteamento em escala 1:5.000, conforme descrito por Sarmento et al. (2008). Apesar de tratar-se de um levantamento convencional, o mapa foi elaborado com o apoio de receptores GPS (*Global Positioning System*) e Sistemas de Informação Geográfica (SIG), encontrando-se disponível em meio digital e georreferenciado, o que facilita seu uso e integração com outras informações espaciais da mesma área.

A Figura 13 mostra o layout do mapa convencional de solos disponível para a área de estudo, elaborado conforme o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006). Observando-se a Figura 13 constata-se que existe um grande número de polígonos de solos representados com cores idênticas. Isso ocorre porque as cores são atribuídas a ordens de solos em nível de ordem, mantendo-se as unidades de mapeamento pertencentes a

uma determinada ordem com a mesma cor. A distinção entre unidades de mapeamento de solos é feita apenas pela inserção da respectiva simbologia no interior de cada polígono.

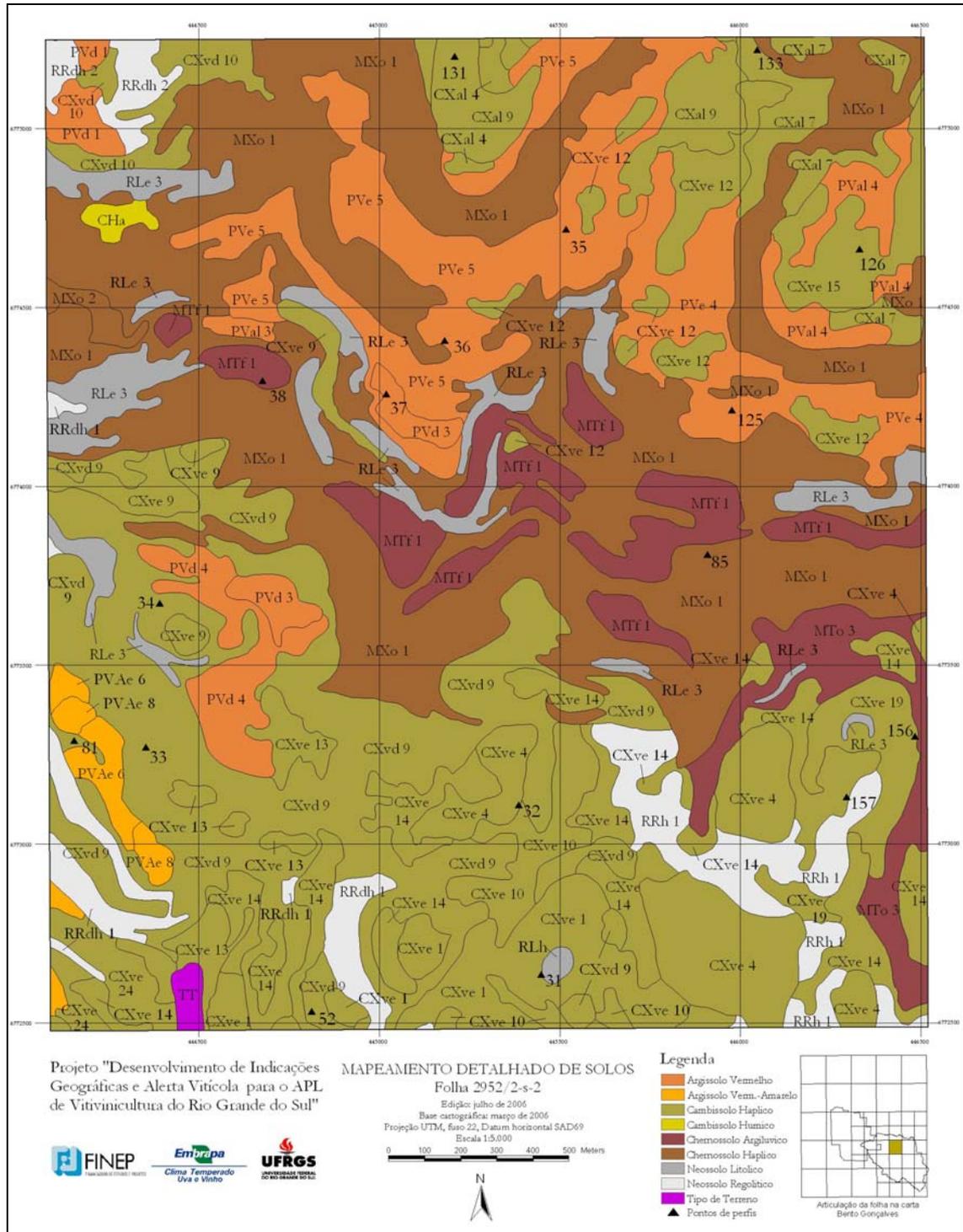


Figura 13. Mapa convencional de solos disponível para a área de estudo (Sarmiento et al., 2008).

O mapa convencional de solos contém 155 polígonos distribuídos em 36 unidades de mapeamento, sendo 10 unidades de Argissolos (15,08% da área), 16 de Cambissolos (41,77%), 4 de Chernossolos (33,96%) e 6 de

Neossolos (9,19%). A Tabela 3 relaciona as 36 unidades de mapeamento de solos identificadas na área de estudo, com a respectiva descrição (Flores, 2006c), número de manchas mapeadas, área ocupada em hectares e proporção da área total. A numeração descontínua deve-se ao fato de que esse mapa é parte de um levantamento maior.

Tabela 3. Sigla, descrição da unidade de mapeamento, número de manchas mapeadas, área em hectares e proporção ocupada por unidade de mapeamento de solos na área de estudo (Flores, 2006c).

Sigla	Descrição da unidade de mapeamento	Nº de manchas	Área (ha)	Prop. (%)
PVd 1	ARGISSOLO VERMELHO Distrófico abrupto A proeminente textura média/muito argilosa fase relevo moderadamente ondulado	2	3,5	0,5
PVd 4	ARGISSOLO VERMELHO Distrófico típico A moderado textura argilosa fase relevo moderadamente ondulado	1	10,8	1,6
PVd 3	ARGISSOLO VERMELHO Distrófico típico A moderado textura argilosa fase relevo moderadamente suave ondulado	2	4,7	0,7
PVe 5	ARGISSOLO VERMELHO Eutrófico típico A proeminente textura argilosa/muito argilosa fase pedregosa relevo forte ondulado	3	46,0	6,8
PVe 4	ARGISSOLO VERMELHO Eutrófico típico A proeminente textura média/argilosa fase relevo forte ondulado	1	20,2	3,0
PVal 4	ARGISSOLO VERMELHO-AMARELO Alítico típico A proeminente textura média/muito argilosa fase pedregosa relevo ondulado	3	8,5	1,3
PVal 3	ARGISSOLO VERMELHO-AMARELO Alítico típico A proeminente textura média/muito argilosa fase relevo suave ondulado	1	1,0	0,1
PVAe 8	ARGISSOLO VERMELHO-AMARELO Eutrófico típico A moderado textura média/argilosa fase relevo moderadamente ondulado	2	2,1	0,3
PVAe 6	ARGISSOLO VERMELHO-AMARELO Eutrófico típico A proeminente textura média fase relevo ondulado	2	4,0	0,6
PVAe 5	ARGISSOLO VERMELHO-AMARELO Eutrófico típico A proeminente textura média/argilosa fase pedregosa relevo ondulado	2	0,9	0,1

Tabela 3. Continuação...

Sigla	Descrição da unidade de mapeamento	Nº de manchas	Área (ha)	Prop. (%)
CXal 9	CAMBISSOLO HÁPLICO Alítico típico A proeminente textura argilosa fase relevo suave ondulado	3	14,0	2,1
CXal 7	CAMBISSOLO HÁPLICO Alítico típico A proeminente textura média com cascalho/argilosa fase pedregosa relevo forte ondulado	7	11,1	1,7
CXal 4	CAMBISSOLO HÁPLICO Alítico típico A proeminente textura média/argilosa fase relevo ondulado	3	1,6	0,2
CXvd 9	CAMBISSOLO HÁPLICO Ta Distrófico úmbrico textura média cascalhenta fase pedregosa relevo forte ondulado	8	94,1	14,0
CXvd 10	CAMBISSOLO HÁPLICO Ta Distrófico úmbrico textura média cascalhenta/argilosa fase pedregosa e rochosa relevo forte ondulado	5	8,2	1,2
CXve 1	CAMBISSOLO HÁPLICO Ta Eutrófico léptico A proeminente textura média fase relevo moderadamente ondulado	6	20,2	3,0
CXve 14	CAMBISSOLO HÁPLICO Ta Eutrófico típico A moderado textura média cascalhenta fase relevo ondulado	18	39,2	5,8
CXve 9	CAMBISSOLO HÁPLICO Ta Eutrófico típico A moderado textura média fase pedregosa relevo ondulado	4	7,7	1,2
CXve 12	CAMBISSOLO HÁPLICO Ta Eutrófico típico A moderado textura média fase relevo ondulado	9	13,2	2,0
CXve 4	CAMBISSOLO HÁPLICO Ta Eutrófico típico A proeminente textura média cascalhenta fase pedregosa relevo suave ondulado	5	25,9	3,8
CXve 19	CAMBISSOLO HÁPLICO Ta Eutrófico típico A proeminente textura média fase pedregosa e rochosa relevo forte ondulado	2	9,7	1,4
CXve 24	CAMBISSOLO HÁPLICO Ta Eutrófico típico A proeminente textura média fase pedregosa e rochosa relevo forte ondulado	2	6,3	0,9
CXve 10	CAMBISSOLO HÁPLICO Ta Eutrófico típico A proeminente textura média fase relevo ondulado	5	10,6	1,6

Tabela 3. Continuação...

Sigla	Descrição da unidade de mapeamento	Nº de manchas	Área (ha)	Prop. (%)
CXve 15	CAMBISSOLO HÁPLICO Ta Eutrófico típico A proeminente textura média/argilosa fase relevo moderadamente ondulado	2	10,2	1,5
CXve 13	CAMBISSOLO HÁPLICO Ta Eutrófico típico textura média cascalhenta fase pedregosa relevo ondulado	6	7,7	1,2
CHa	CAMBISSOLO HÚMICO Alumínico típico textura argilosa/muito argilosa fase relevo suave ondulado	1	1,5	0,2
MTf 1	CHERNOSSOLO ARGILÚVICO Férrico típico textura argilosa/muito argilosa fase pedregosa relevo forte ondulado	9	29,3	4,4
MTo 3	CHERNOSSOLO ARGILÚVICO Órtico típico textura argilosa cascalhenta/muito argilosa cascalhenta fase epipedregosa relevo forte ondulado	1	15,9	2,4
MXo 1	CHERNOSSOLO HÁPLICO Órtico léptico textura média cascalhenta fase pedregosa e rochosa relevo montanhoso	8	181,7	27,0
MXo 2	CHERNOSSOLO HÁPLICO Órtico típico textura média fase pedregosa relevo forte ondulado	1	2,0	0,3
RLe 3	NEOSSOLO LITÓLICO Eutrófico A moderado textura média fase pedregosa relevo montanhoso e escarpado	16	26,7	4,0
RLh	NEOSSOLO LITÓLICO Húmico típico textura média fase pedregosa relevo ondulado	1	0,7	0,1
RRdh 1	NEOSSOLO REGOLÍTICO Distro-úmbrico léptico textura média fase pedregosa relevo montanhoso	6	11,1	1,6
RRdh 2	NEOSSOLO REGOLÍTICO Distro-úmbrico típico textura argilosa cascalhenta fase pedregosa relevo ondulado	2	4,8	0,7
RRe 4	NEOSSOLO REGOLÍTICO Eutrófico típico A proeminente textura média fase pedregosa e rochosa relevo forte ondulado	1	1,5	0,2
RRh 1	NEOSSOLO REGOLÍTICO Húmico léptico textura média cascalhenta fase relevo suave ondulado	5	17,0	2,5
Total		155	673,5	100,0

3.2.2. Base cartográfica digital

A base cartográfica digital disponível para a área de estudo foi produzida por empresa especializada em levantamentos aerofotogramétricos, com financiamento da FINEP (Financiadora de Estudos e Projetos) através da Encomenda Ação Transversal Arranjos Produtivos Locais 01/2004, conhecida como “Projeto APL Vinhos”. O projeto é coordenado pela Embrapa Uva e Vinho e tem como parceiros a Embrapa Clima Temperado, a Universidade de Caxias do Sul (UCS) e a Universidade Federal do Rio Grande do Sul, através do Centro de Ecologia (UFRGS/CENECO).

Conforme relatório técnico do aerolevanteamento, as tomadas das fotografias se deram no dia 01/11/2005, sob céu sem nuvens, fumaça ou bruma. A restituição fotogramétrica, correspondente à fotointerpretação das imagens aéreas a partir dos estereomodelos de pares de fotos digitais, foi realizada no período de 15/11/2005 a 10/03/2006. O processo empregou técnicas de fotogrametria digital para a obtenção de um mapeamento planialtimétrico seguindo as especificações técnicas para a escala 1:5.000, o qual cobre aproximadamente 170 km² e abrange integralmente o Vale dos Vinhedos e o município de Monte Belo do Sul.

O produto final do aerolevanteamento inclui um conjunto de fotografias digitais ortorretificadas individuais, um mosaico digital contínuo das fotografias ortorretificadas, com resolução espacial de 0,6 metros, e um conjunto de arquivos gráficos vetoriais em formato de software CAD (*Computer Aided Design*), organizados em folhas, contendo as principais feições cartográficas da planimetria (cursos e corpos d’água, rodovias, ferrovias, estradas vicinais, etc.) e da altimetria (curvas de nível com equidistância de 5 metros e pontos cotados), além de vários elementos auxiliares (grade de coordenadas, escala, indicação de Norte, indicação da folha, etc.).

Para o aproveitamento do material em ambiente de SIG foi necessária uma etapa de estruturação da base cartográfica digital, de forma a possibilitar a realização das análises necessárias e a geração das informações de interesse, conforme descrito por Sarmiento et al. (2008). Esta etapa envolveu a seleção e extração dos *layers* (camadas) úteis de cada folha do aerolevanteamento, sua concatenação e a edição das bordas das folhas para a união dos elementos. Como resultado, obteve-se um conjunto de planos de

informação individuais e contínuos (Figura 14), topologicamente estruturados e vinculados a tabelas com os principais atributos dos elementos.

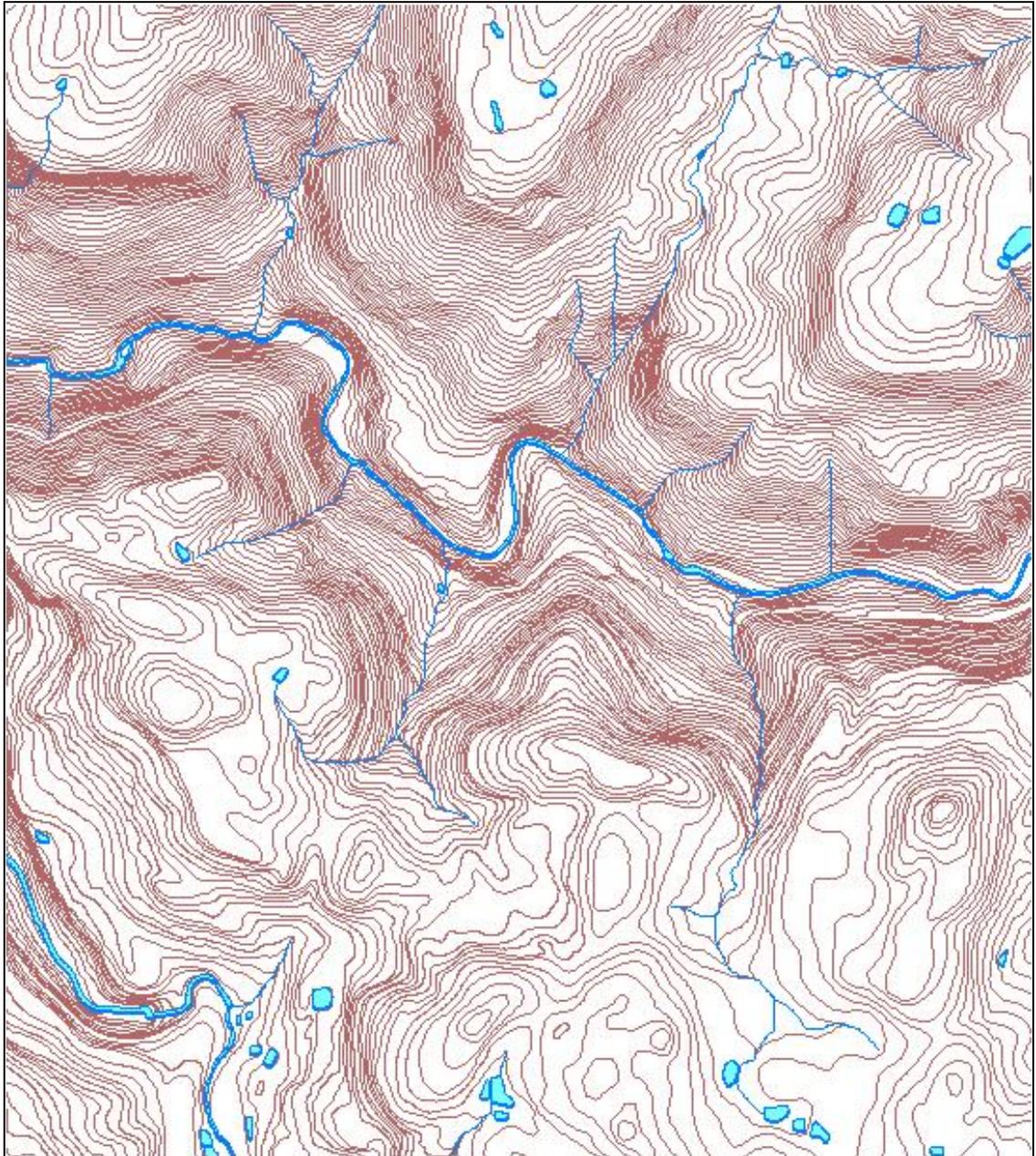


Figura 14. Rede de drenagem superficial e curvas de nível com equidistância vertical de 5 metros, extraídas da base cartográfica digital disponível para a área de estudo (Sarmiento et al., 2008).

3.2.3. Modelo Numérico do Terreno

O Modelo Numérico do Terreno (MNT) disponível para a área de estudo foi gerado a partir do plano de informação contendo as curvas de nível com equidistância vertical de 5 metros, através de interpolação linear baseada em uma rede triangular irregular (*Triangulated Irregular Network* - TIN). Conforme descrito por Sarmiento et al. (2008), para melhorar a qualidade do

resultado, empregou-se uma função parabólica para ajustar as quebras de relevo e eliminar os efeitos de “ponte” e de “túnel” (estimar a altitude do fundo dos vales e do topo de elevações). A opção pela interpolação TIN levou em consideração as vantagens que esse tipo de modelo introduz na representação de áreas com relevo complexo (Câmara & Medeiros, 1996). Redes triangulares podem facilmente incorporar descontinuidades no terreno e são estruturas eficientes porque a densidade dos triângulos pode variar para se ajustar à rugosidade do terreno (Moore et al., 1991; Wilson & Gallant, 2000).

Para facilitar o uso posterior do MNT no cálculo de atributos do terreno (variáveis topográficas e hidrológicas), os valores de elevação derivados da rede triangular irregular foram interpolados e convertidos para o modelo *raster* (matricial) adotando-se uma resolução espacial de cinco (5) metros (célula de 5 m x 5 m). A Figura 15 mostra uma visualização em perspectiva do MNT da área de estudo.

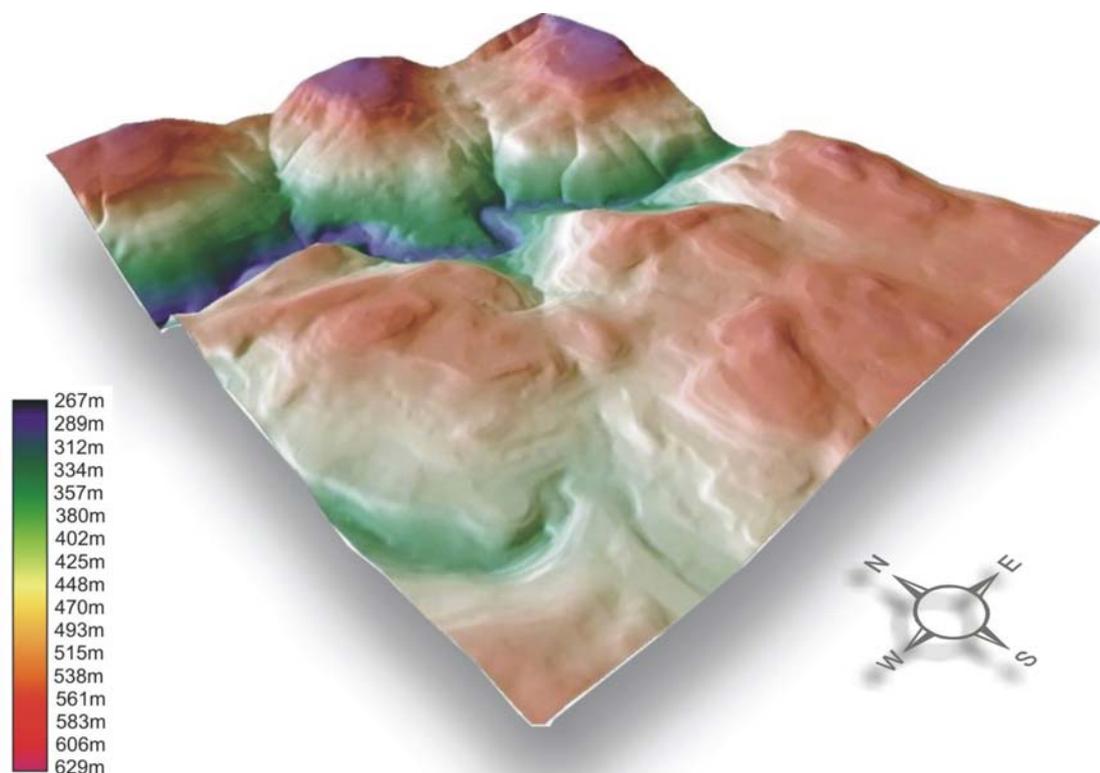


Figura 15. Visualização em perspectiva do MNT da área de estudo, vista de sudoeste para nordeste.

3.2.4. Softwares

Um número relativamente grande de *softwares* tem sido utilizado nos estudos de mapeamento digital de solos, mas nenhum deles é capaz de satisfazer todas as necessidades de análise. O problema é que cada pacote é especializado em um determinado conjunto de operações, estatísticas ou espaciais, e não há muitos que suportem plenamente ambas. Por essa razão, freqüentemente torna-se necessário lançar mão de vários *softwares* para levar a cabo um trabalho de mapeamento digital de solos.

Hengl & Heuvelink (2004), enfocando os desafios para o mapeamento preditivo de solos, afirmam que uma limitação importante é que essas ferramentas não são facilmente utilizáveis por não-especialistas. Embora muitos métodos sejam testados e documentados, ainda há uma grande lacuna entre o que é possível para alguns e o que está disponível para muitos. Os autores compararam 10 *softwares* estatísticos e de SIG amplamente disseminados e concluíram que somente o Idrisi reúne, de forma integrada, um amplo conjunto de operações espaciais e estatísticas sofisticadas, importantes para modelagem e predição robusta. Rossiter (2008) também o destaca como ferramenta analítica robusta e de baixo custo.

Entretanto, esse *software* tem sido pouco utilizado em trabalhos de mapeamento digital de solos. Neste estudo empregou-se o *software* de SIG Idrisi Taiga, versão 16.02 (Eastman, 2009b) para a execução de todos os cálculos e análises. A UFRGS possui uma licença institucional do Idrisi, com permissão de acesso e uso ilimitado por pesquisadores e estudantes.

3.3. Metodologia

A Figura 16 representa esquematicamente, através de um fluxograma, os principais passos metodológicos empregados neste estudo.

O primeiro passo foi o agrupamento do mapa convencional de solos para o primeiro nível categórico do Sistema Brasileiro de Classificação de Solos (Embrapa, 2006). No segundo passo, foram gerados os pontos amostrais, utilizando-se cinco densidades de amostragem dentro do intervalo recomendado para levantamentos detalhados de solos, que varia de 0,2 a 4 observações por hectare (IBGE, 2007). O terceiro passo envolveu a geração do conjunto de variáveis a serem utilizadas como preditoras das ordens de solos, em um total de sete, sendo seis derivadas do MNT e uma derivada da rede de

drenagem superficial. O quarto passo envolveu o cruzamento dos pontos de amostragem com as ordens de solos e com as variáveis predictoras, para o treinamento dos algoritmos de classificação, e a predição de ocorrência das ordens de solos. O último passo envolveu a avaliação de acurácia, através do cruzamento entre os mapas estimados e o mapa original de ordens de solos, obtendo-se três medidas quantitativas sobre o desempenho de cada algoritmo de classificação com as diferentes densidades de amostragem. Cada um dos passos acima é descrito em maior detalhe a seguir.

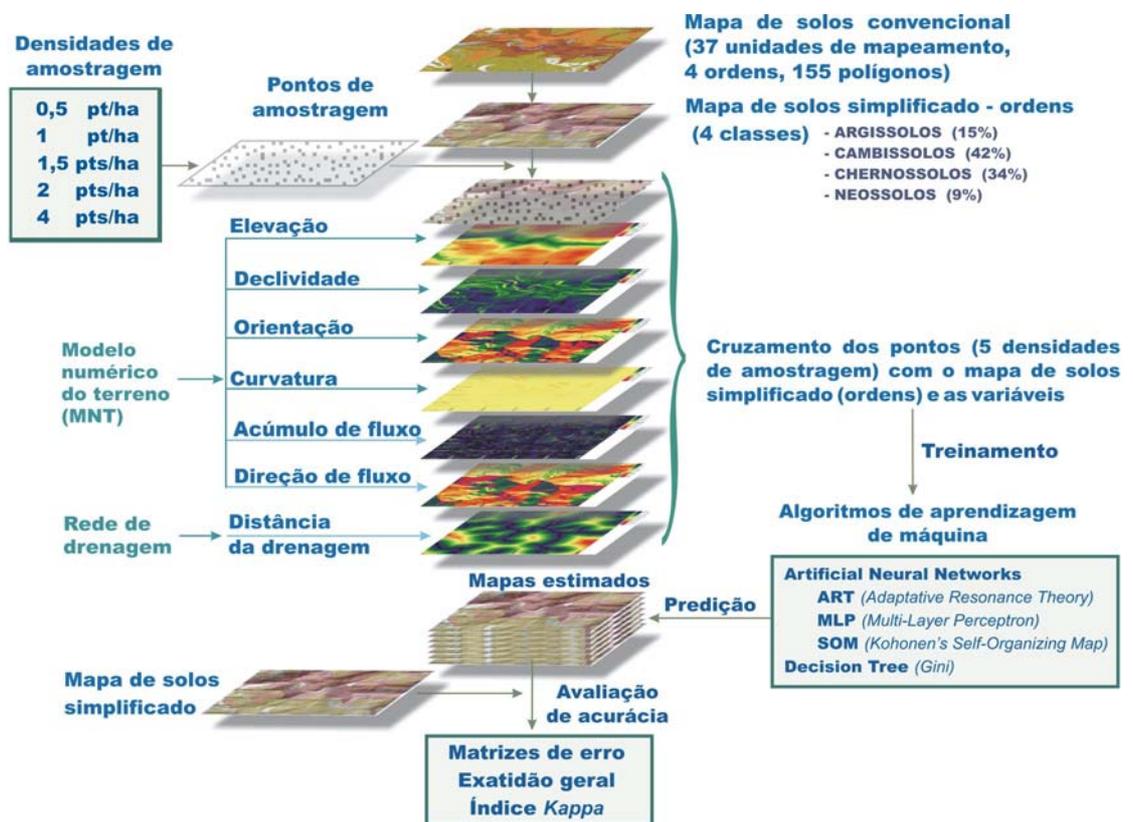


Figura 16. Fluxograma com os principais passos da metodologia utilizada.

3.3.1. Agrupamento do mapa de solos

A classificação em nível de unidade de mapeamento é baseada em um grande número de características físicas e químicas, obtidas através de observações diretas de perfis de solo e de amostras coletadas em campo e analisadas em laboratório. Como neste estudo pretendeu-se estimar informações de solos a partir de variáveis relacionadas a atributos do terreno, derivadas de dados cartográficos, o mapa convencional de solos disponível

necessitou ser simplificado para um nível de detalhe temático coerente com essas variáveis.

A simplificação do mapa foi feita através do agrupamento das unidades de mapeamento individuais em classes mais genéricas, com o objetivo de reduzir a complexidade do mapa convencional de solos e dar maior consistência a essas classes. As 36 unidades de mapeamento do mapa original foram agrupadas para o primeiro nível do Sistema Brasileiro de Classificação de Solos (Embrapa, 2006), através da função RECLASS do Idrisi, resultando em um mapa de 4 classes, correspondentes às ordens de solos.

Segundo Santos et al. (2005), as características diferenciais para os níveis categóricos mais elevados da classificação de solos devem ser propriedades dos solos que resultam diretamente dos processos de gênese do solo ou que afetam, diretamente, a gênese do mesmo, porque estas propriedades apresentam um maior número de características acessórias. As diversas classes no 1º nível categórico são separadas pela presença ou ausência de determinados atributos, horizontes diagnósticos ou propriedades que são passíveis de serem identificadas no campo mostrando diferenças no tipo e grau de desenvolvimento dos processos que atuaram na formação do solo. Assim, como o relevo é um fator de formação preponderante na área de estudo, é esperado que a distribuição das ordens de solos possa ser estimada a partir de variáveis relacionadas a atributos do terreno.

A Tabela 4 relaciona a superfície em hectares e a proporção da área de estudo ocupada pelas ordens de solos no mapa simplificado e a Figura 17 mostra o mapa original e o mapa simplificado, resultante do agrupamento. O mapa de solos simplificado constituiu a informação de referência para o estabelecimento das relações entre as variáveis preditoras e a distribuição espacial dos solos, a partir de um conjunto de pontos de amostragem aleatoriamente distribuídos na área de estudo.

Tabela 4. Área em hectares e proporção ocupada na área de estudo pelos solos agrupados para o primeiro nível (Ordem).

Ordem de solos	Área ocupada (ha)	Proporção (%)
Argissolos	101,6	15,08
Cambissolos	281,3	41,77
Chernossolos	228,7	33,96
Neossolos	61,9	9,19
Total	673,5	100,00

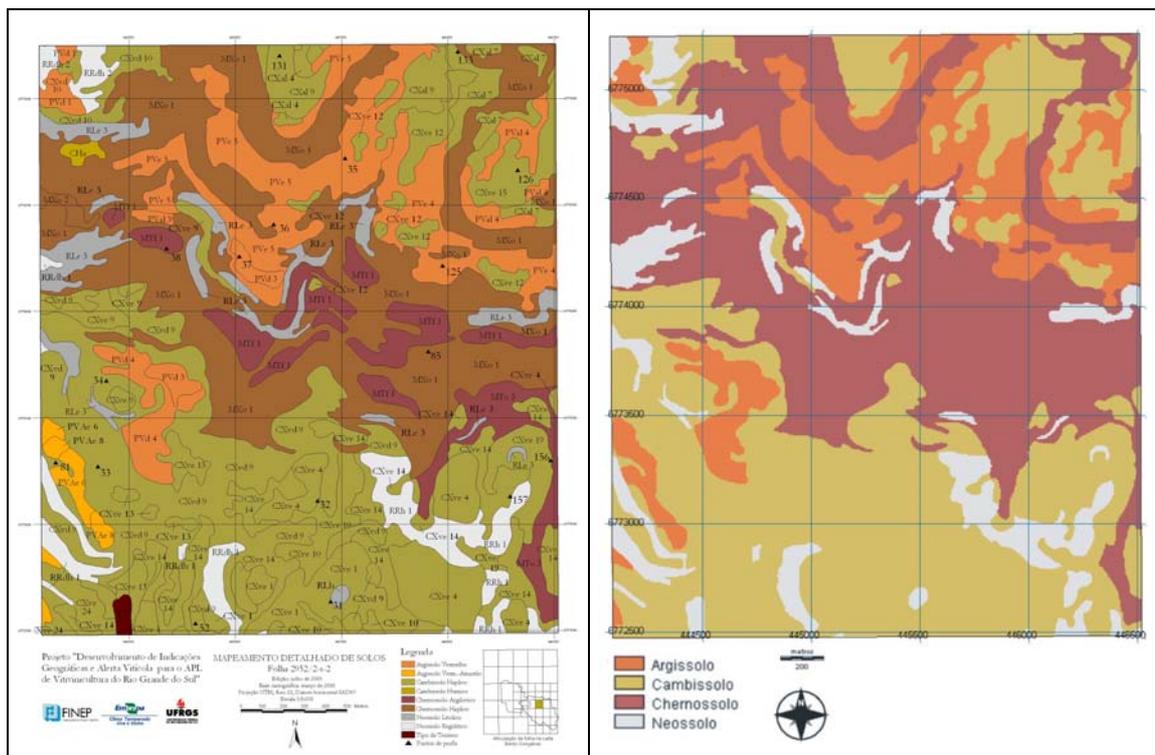


Figura 17. Mapa convencional de solos, com 36 unidades de mapeamento, e mapa agrupado para o primeiro nível (Ordem), com 4 classes.

3.3.2. Geração dos pontos de amostragem

A geração dos pontos de amostragem foi feita com a rotina SAMPLE, própria para amostragem espacial, que produz como resultado um arquivo vetorial de pontos. O esquema de amostragem adotado foi completamente aleatório, ou seja, uma distribuição aleatória dos pontos na superfície da área de estudo. Com vistas a avaliar a sensibilidade dos algoritmos de classificação à densidade de amostragem foram gerados conjuntos de pontos em cinco diferentes densidades de amostragem, dentro do intervalo recomendado para levantamentos detalhados de solos (IBGE, 2007).

As densidades utilizadas foram de 0,5 ponto por hectare (1 ponto a cada 2 hectares), 1 ponto por hectare, 1,5 pontos por hectare (3 pontos a cada 2 hectares), 2 pontos por hectare e 4 pontos por hectare. Os arquivos vetoriais de pontos foram convertidos para o modelo *raster*, através da rotina RASTERVECTOR, e os arquivos *raster* resultantes usados para identificar a ordem de solos correspondente à posição de cada ponto, usando-se a rotina EXTRACT, tendo o mapa simplificado como referência. A relação entre cada ponto e a respectiva ordem de solo foi salva em um arquivo de valores (*values file*), uma lista de duas colunas em que a primeira contém o número do ponto e a segunda a ordem de solo. Essa lista foi então utilizada para classificar cada ponto na respectiva ordem de solo, com a rotina ASSIGN.

Como resultado obteve-se cinco mapas *raster*, um para cada densidade de amostragem, onde os pontos encontram-se classificados com o código da ordem de solo correspondente à sua posição. A Tabela 5 relaciona as cinco densidades de amostragem (pontos.ha⁻¹), o número total de pontos e a distribuição do número de pontos por ordem de solo em cada densidade. A Figura 18 mostra o mapa de solos simplificado seguido da sobreposição dos pontos referentes às cinco densidades de amostragem, ilustrando a sua distribuição espacial.

Tabela 5. Densidade de amostragem, número total de pontos e número de pontos por ordem de solos.

Densidade (pontos.ha ⁻¹)	Nº de pontos	Nº de pontos por ordem de solos			
		Argissolos	Cambissolos	Chernossolos	Neossolos
0,5	322	45	145	105	27
1	644	113	278	204	49
1,5	966	166	396	320	84
2	1.288	209	544	419	116
4	2.576	419	1036	888	233

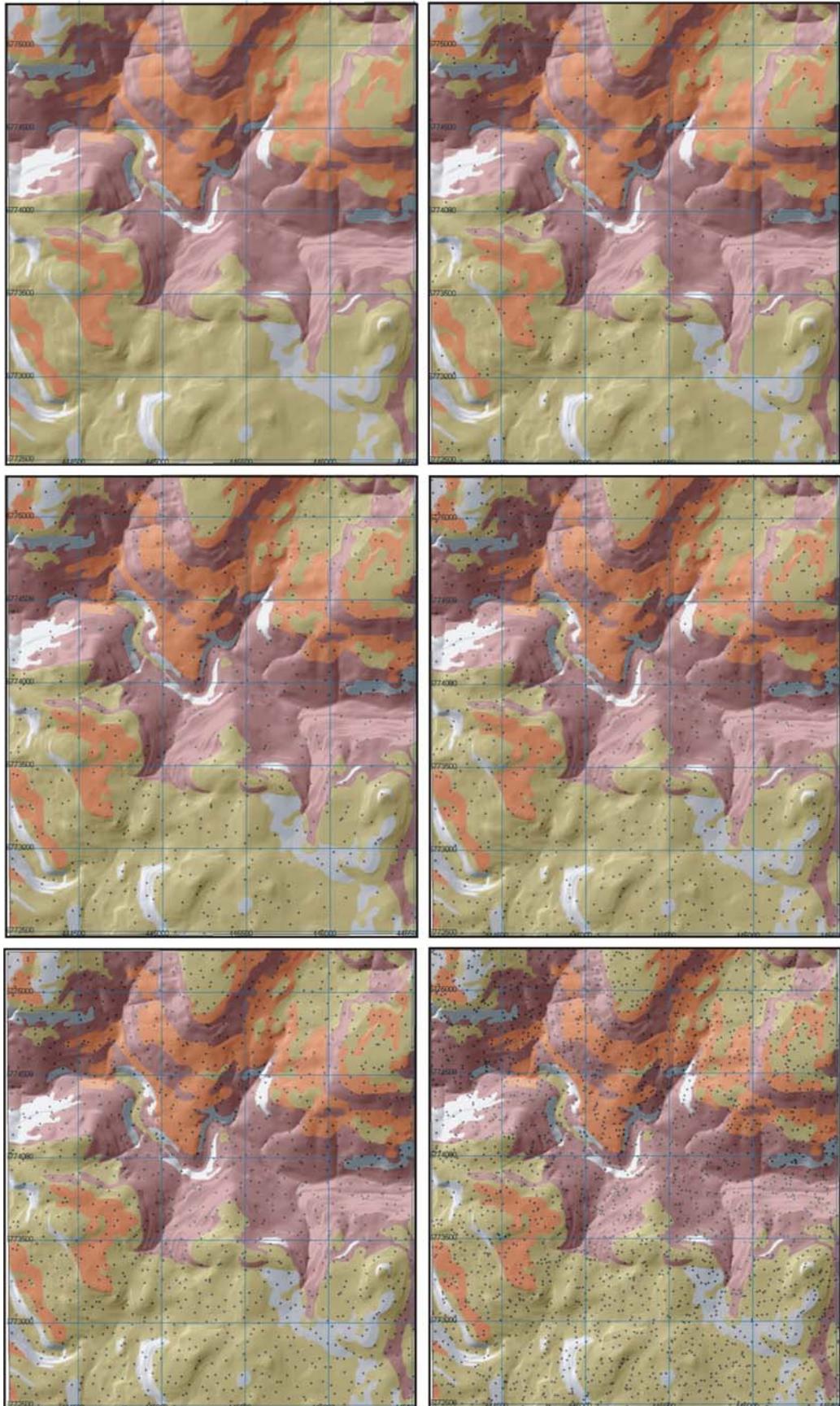


Figura 18. Mapa de solos simplificado sobreposto com os pontos das cinco densidades de amostragem.

3.3.3. Cálculo das variáveis preditoras

Como o relevo é um dos fatores de formação dos solos e a área de estudo apresenta topografia movimentada, é esperado que variáveis descritoras do relevo possam explicar a ocorrência de ordens de solos. Além da altitude (MNT), foram calculadas três variáveis topográficas propriamente ditas (declividade, orientação e curvatura) e duas variáveis hidrológicas, decorrentes da topografia (direção de fluxo e acúmulo de fluxo). A partir da planimetria, constante na base cartográfica, foi calculada ainda a distância da rede de drenagem superficial.

Estas variáveis descrevem importantes aspectos relacionados à gênese do solo e à distribuição de sua ocorrência na paisagem, tais como regime hídrico, erosão e deposição de sedimentos, escarpamento de morros, concentração de matéria orgânica, profundidade do horizonte A, entre outros (Florinsky et al., 2002).

3.3.3.1. Elevação (MNT)

A elevação expressa a energia potencial do relevo e está também relacionada a gradientes climáticos locais (Wilson & Gallant, 2000; Romano & Chirico, 2004). É usualmente representada através de um MNT, arquivo matricial (*raster*) onde o valor armazenado em cada célula (*pixel*) representa a altitude do terreno.

A elevação da área de estudo varia de 267 metros a 629 metros acima do nível do mar, com uma distribuição espacial heterogênea devido ao relevo movimentado (Figura 19). Os locais de menor elevação situam-se junto no vale do arroio Leopoldina e os mais altos correspondem ao topo de algumas elevações situadas na porção norte e nordeste da área.

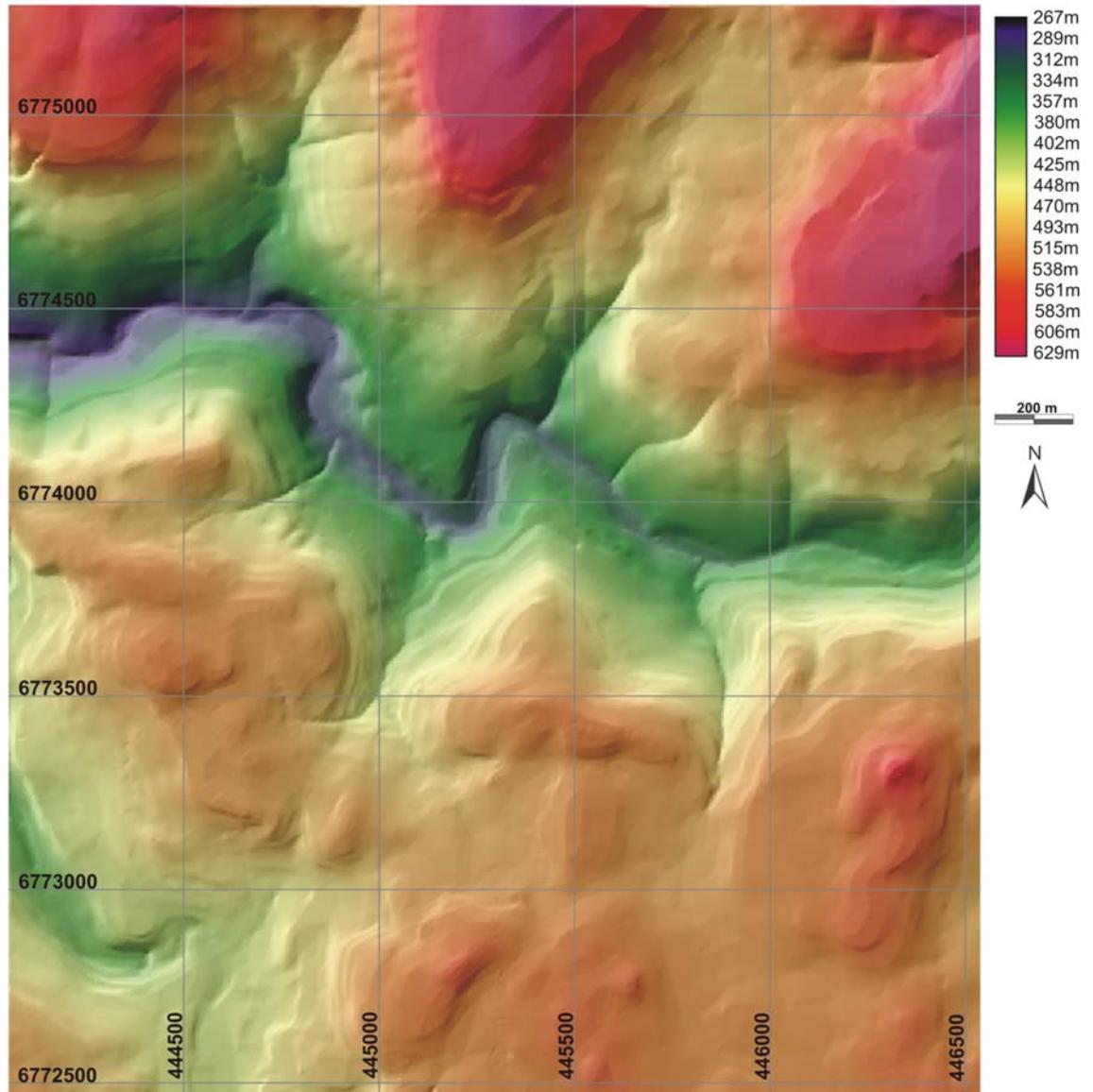


Figura 19. Mapa de elevação (MNT) da área de estudo.

3.3.3.2. Declividade

A declividade foi calculada com a rotina SURFACE. Segundo Eastman (2009a), essa rotina determina a declividade de uma célula com base na sua resolução, na sua altitude e na altitude das células imediatamente adjacentes ortogonalmente à mesma: ao topo (norte), à base (sul), à esquerda (oeste) e à direita (leste). A declividade é calculada como o vetor resultante da declividade ao longo do eixo X e da declividade ao longo do eixo Y, conforme a Equação 7 abaixo:

$$\tan_dec = \sqrt{\left(\frac{\text{esquerda} - \text{direita}}{\text{res} \times 2}\right)^2 + \left(\frac{\text{superior} - \text{inferior}}{\text{res} \times 2}\right)^2} \quad (7)$$

onde \tan_{dec} é a tangente do ângulo que apresenta a máxima inclinação no sentido da pendente; *esquerda*, *direita*, *superior*, *inferior* são os valores das células imediatamente adjacentes e *res* é a resolução da célula. O valor de \tan_{dec} multiplicado por 100 produz o resultado em porcentagem, enquanto o arco tangente de \tan_{dec} produz o resultado em graus. Do ponto de vista matemático, a declividade representa a primeira derivada da altitude em função da distância horizontal.

Neste estudo a declividade foi calculada em porcentagem. Conforme mostra a Figura 20, a declividade na área de estudo varia de 0,11%, em locais praticamente planos situados em alguns patamares e em topos de elevações, até 237,31%, em paredões íngremes encontrados no vale do arroio Leopoldina e alguns outros locais.

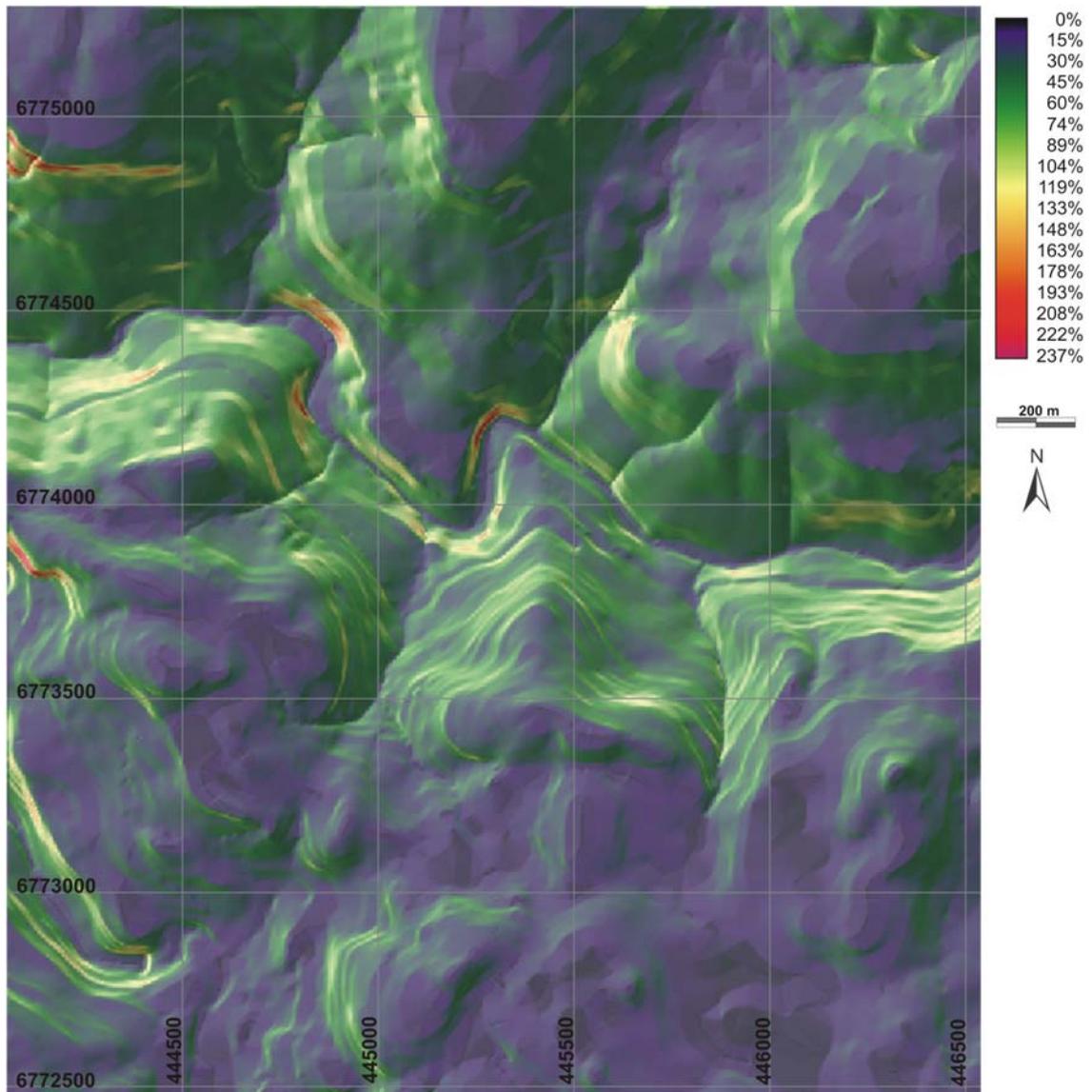


Figura 20. Mapa de declividade da área de estudo, expressa em porcentagem.

3.3.3.3. Orientação

A orientação foi calculada com o módulo SURFACE, com base nos mesmos elementos da Equação 7 anteriormente descrita. Em áreas onde a declividade é perfeitamente plana (declividade = 0%) não existe propriamente uma orientação da superfície. Nestes locais é utilizado o valor -1 , com vista a diferenciá-los das demais áreas que possuem realmente uma orientação.

Conforme mostra a Figura 21, há uma grande variação na orientação das encostas na área de estudo, conseqüência do relevo movimentado. É possível visualizar os fundos de vales correspondentes à rede de drenagem superficial e os divisores de água, ambos representados por cores contrastantes, pois ocorrem entre células de orientações opostas.

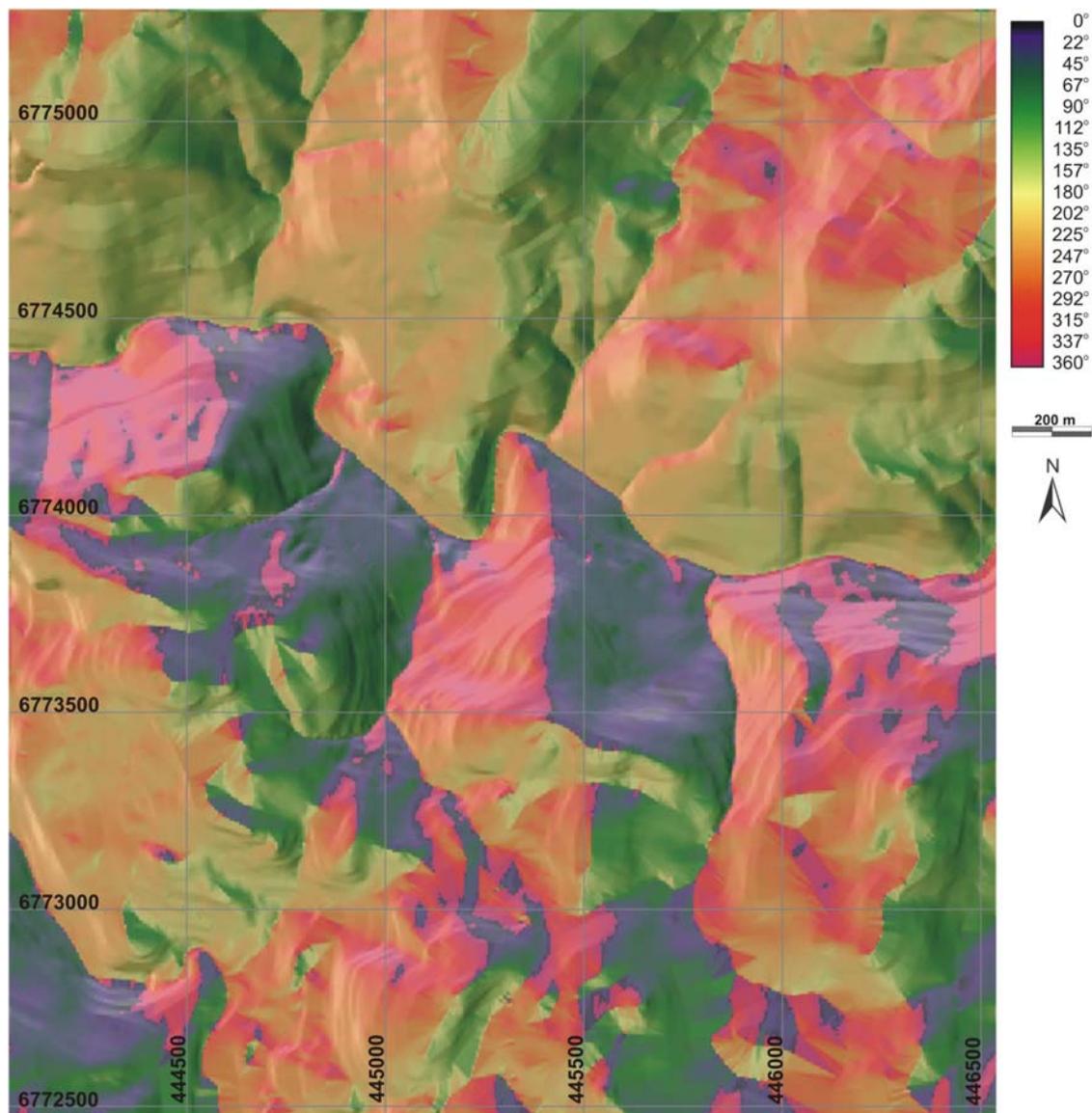


Figura 21. Mapa de orientação das encostas na área de estudo.

3.3.3.4. Curvatura

Neste estudo foi calculada a curvatura de perfil (ou curvatura máxima), empregando-se a rotina CURVATURE. Os cálculos da curvatura baseiam-se no ajuste de uma superfície polinomial sobre uma área de 3 x 3 células, obtendo para a célula central a magnitude da taxa de mudança de uma linha tangente à curva matematicamente ajustada na direção da sua orientação. Os valores são positivos quando a superfície é convexa, pois a taxa de variação da declividade aumenta ao longo da encosta, e negativos quando a superfície é côncava, pois a taxa de variação da declividade diminui ao longo da encosta, (Figura 22).

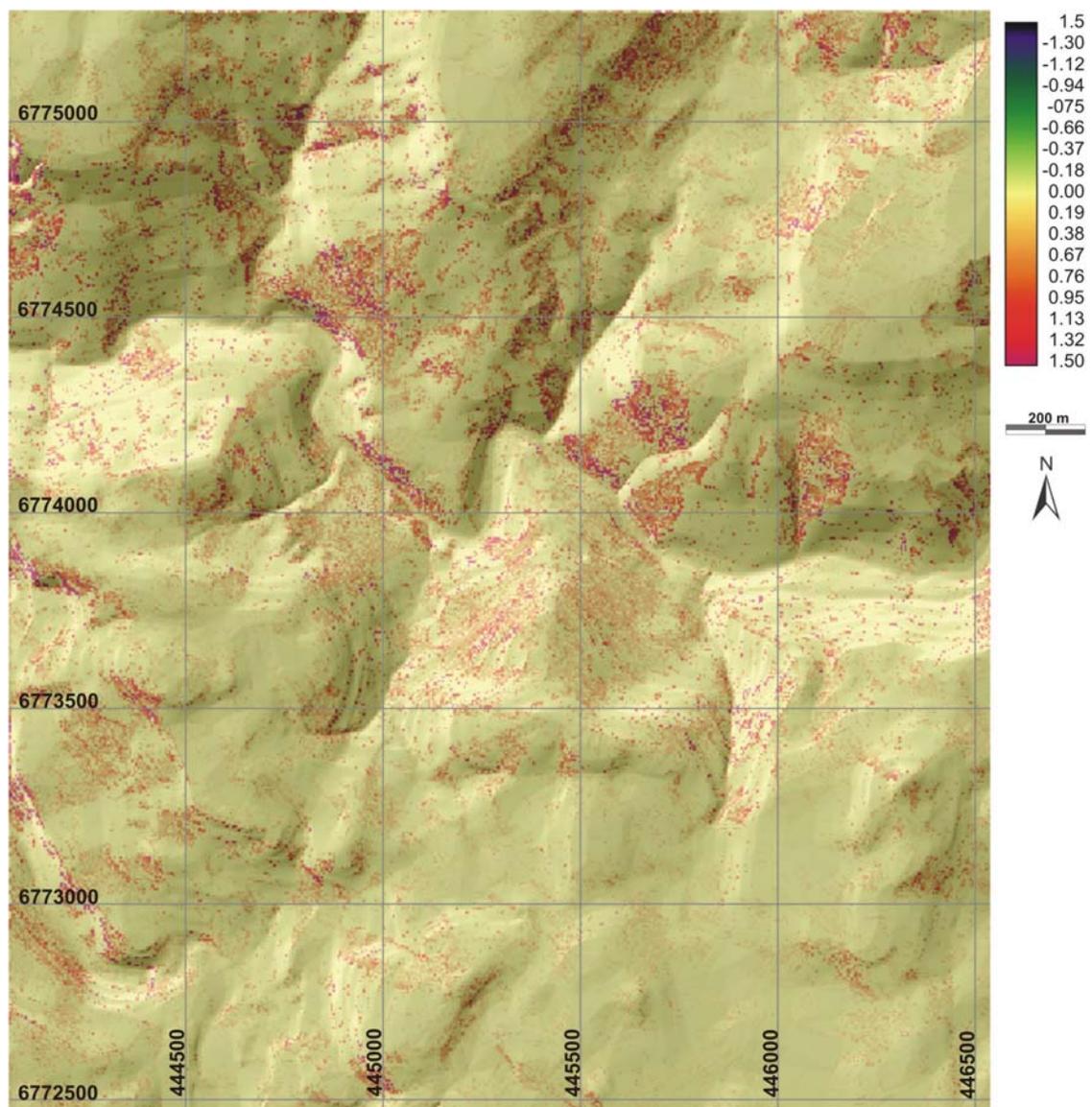


Figura 22. Mapa de curvatura da área de estudo.

Conforme mostra o mapa da Figura 22, a maior parte da área apresenta valores da curvatura de perfil em torno de zero, o que significa

encostas com declividade aproximadamente constante. Os valores mais afastados de zero estão relacionados principalmente à inflexão do relevo junto às bordas dos derrames basálticos, convexos no topo e côncavos na base.

3.3.3.5. Direção de fluxo

A direção de fluxo foi calculada com a rotina FLOW. O algoritmo desta rotina examina uma janela de três por três células, comparando a altitude da célula central com a altitude das oito vizinhas, e atribui a direção que corresponde à maior declividade encontrada. No caso de áreas planas, a janela de 3 por 3 células é aumentada a fim de possibilitar a avaliação da forma da superfície além dessa área. A Figura 23 mostra o mapa de direção de fluxo calculado para a área de estudo.

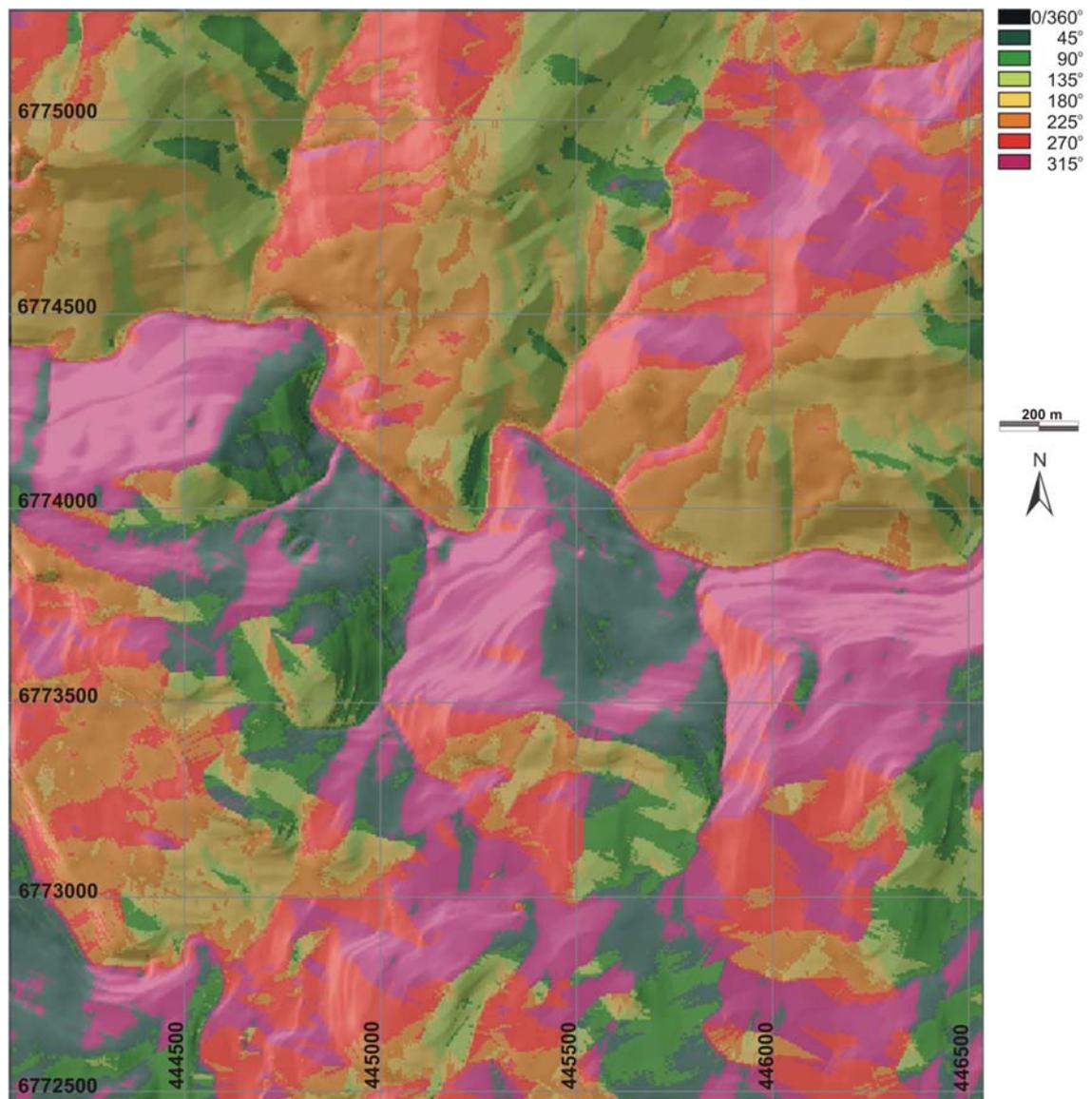


Figura 23. Mapa de direção de fluxo da área de estudo.

Assim como a orientação da superfície, os valores de direção do fluxo variam de 0° a 360° no sentido horário a partir do Norte, mas não são contínuos. Como eles indicam a célula que recebe o fluxo, as direções são limitadas a oito possibilidades, correspondentes às oito células vizinhas de uma determinada célula em análise. Assim, a direção de fluxo varia de 45° em 45°, conforme o quadrante em que se encontra a célula de destino.

3.3.3.6. Acúmulo de fluxo

O acúmulo de fluxo foi calculado com a rotina RUNOFF, que adota o algoritmo descrito por Jenson & Domingue (1988), de amplo uso em SIG. RUNOFF permite levar em conta a precipitação, a taxa de infiltração, a duração da precipitação e da infiltração e os valores iniciais de absorção pela superfície para calcular o volume de escoamento superficial que chega a uma determinada célula do MNT. Neste trabalho foi realizada uma análise mais simples, considerando-se que toda a precipitação que cai em uma célula escoar para a célula situada imediatamente abaixo na encosta, totalizando-se o número de células a montante que potencialmente podem contribuir para o volume de água concentrado em uma determinada célula, ou seja, o tamanho da bacia de contribuição para esta célula.

A Figura 24 mostra o mapa de acúmulo de fluxo, onde a cada célula é atribuído o valor equivalente ao número de células das quais ela recebe fluxo. As células com os maiores valores de fluxo acumulado constituem as linhas de fluxo, que formam a rede de drenagem, representadas por cores escuras no mapa. A célula com o maior acúmulo de fluxo na área de estudo recebe a contribuição de 204.765 células, mas a legenda do mapa foi escalonada para diminuir o contraste visual entre o maior e o menor acúmulo de fluxo. Se uma célula possui o valor zero significa que nenhuma outra célula escoar para ela, de forma que ela constitui um máximo local, correspondendo a um pico ou à linha do divisor de águas (Li et al., 2005). Estas células são representadas por cores claras no mapa.

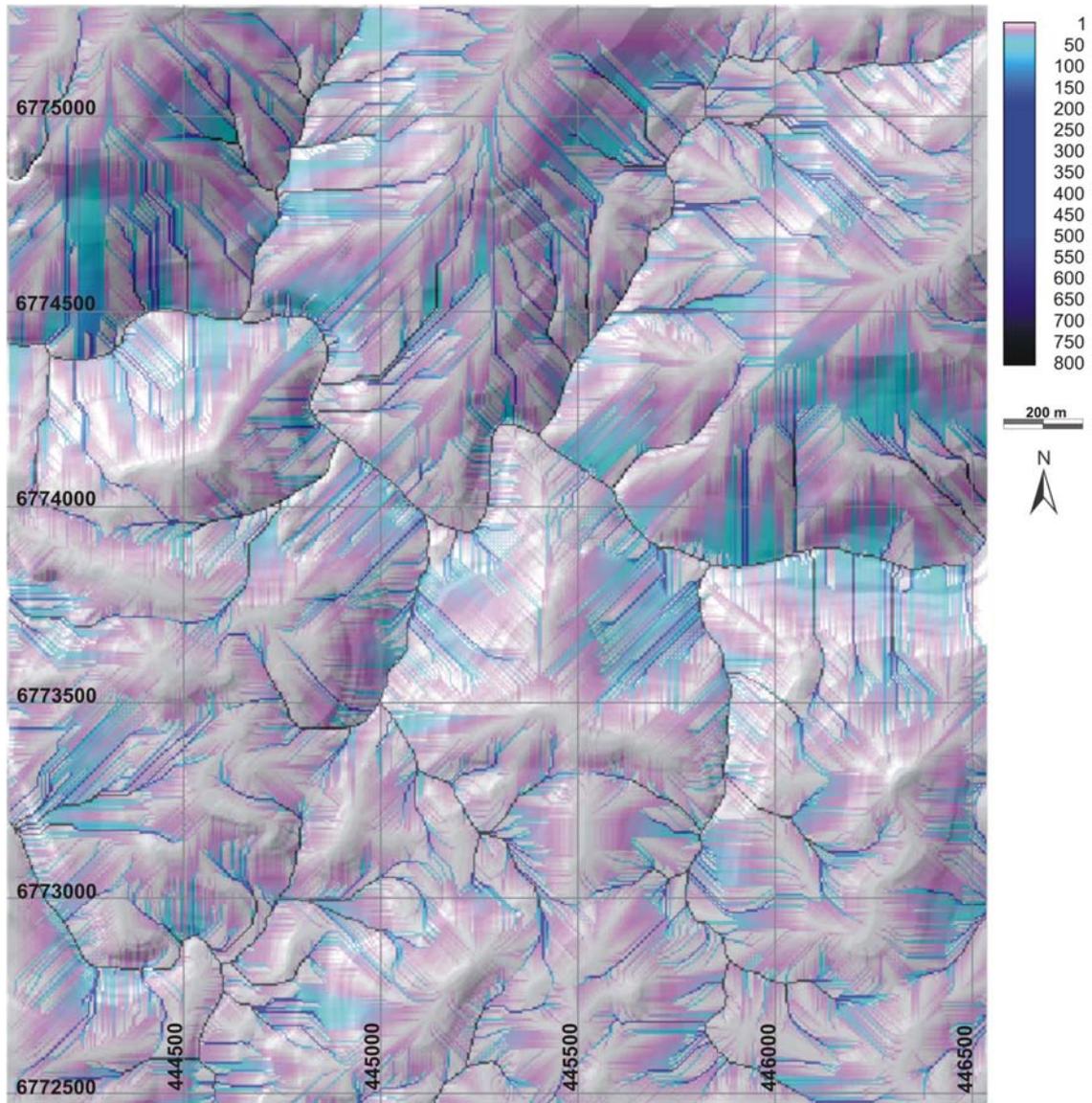


Figura 24. Mapa de acúmulo de fluxo da área de estudo.

3.3.3.7. Distância da rede de drenagem

A distância da rede de drenagem superficial foi calculada a partir dos cursos d'água constantes na base cartográfica digital da área de estudo. O arquivo vetorial da rede de drenagem foi inicialmente convertido para o modelo *raster*, através da rotina RASTERVECTOR, e o cálculo da distância então realizado com a rotina DISTANCE.

O mapa de distâncias obtido (Figura 25) contém em cada célula o valor da sua distância euclidiana, em linha reta no plano horizontal, até a célula de rede de drenagem mais próxima. As distâncias são obtidas nas mesmas unidades do sistema de referência, no presente caso em metros, uma vez que a base cartográfica se encontra na projeção UTM. Conforme mostra a legenda

no mapa de distâncias, o local mais afastado da rede de drenagem na área de estudo encontra-se a uma distância de 608,77 metros.

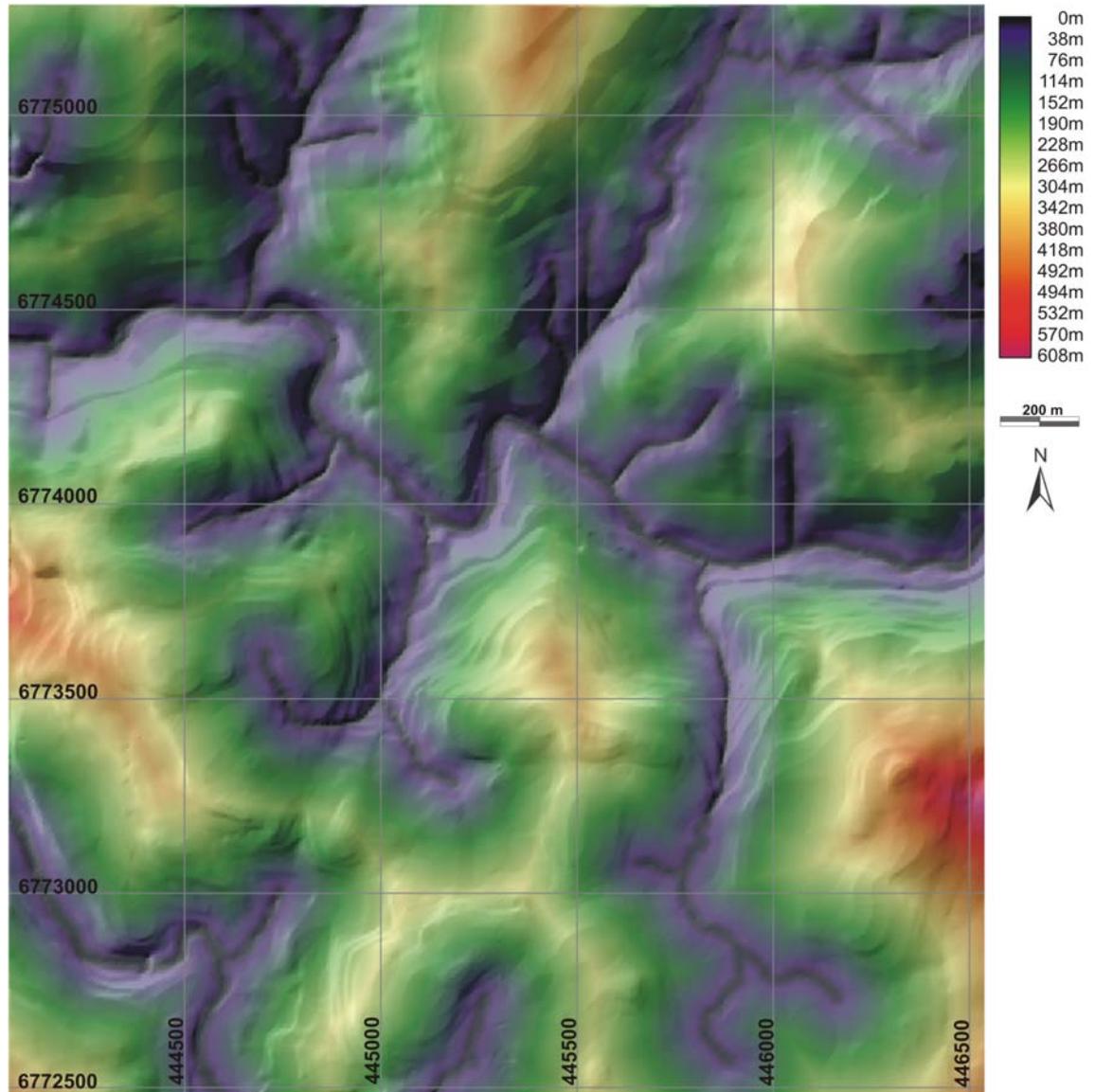


Figura 25. Mapa de distâncias da rede de drenagem na área de estudo.

3.3.4. Treinamento dos algoritmos de classificação e predição da ocorrência das ordens de solos

Para o estabelecimento das relações entre as variáveis anteriormente descritas e a distribuição espacial das ordens de solos foram utilizados quatro algoritmos de classificação baseados no conceito de aprendizagem de máquina, sendo três implementações de redes neurais artificiais e uma árvore de decisão. De acordo com Lippitt et al. (2008), os dois tipos de algoritmos de aprendizagem de máquina são abstrações do processo de aprendizagem humana, mas diferem fundamentalmente na sua abordagem.

Redes neurais simulam a estrutura de neurônios e conexões do cérebro humano, enquanto árvores de decisão simulam o processo de abstração humana através de uma categorização hierárquica.

O treinamento dos algoritmos foi realizado empregando os pontos de amostragem classificados com o código da ordem de solo e os mapas das variáveis. Durante o treinamento, em cada ponto são extraídos os valores de cada variável utilizada, visando o estabelecimento das relações entre as variáveis preditoras e a distribuição espacial das ordens de solos. Todos os algoritmos foram treinados empregando os mesmos conjuntos de pontos amostrais classificados com o código da ordem de solo, em cinco densidades de amostragem. Depois de treinado com os pontos de amostragem, o algoritmo foi então usado para classificar as ordens de solos no restante da área de estudo, a partir dos valores das variáveis preditoras.

3.3.4.1. Rede neural *Multi-layer Perceptron* (MLP)

Redes neurais do tipo *Multi-layer Perceptrons* (MLP) são calibradas através de um algoritmo de retropropagação baseado em um procedimento de aprendizado recursivo que busca minimizar o erro de treinamento da rede. Inicialmente os neurônios da camada de entrada e da(s) camada(s) oculta(s) são ponderados aleatoriamente e cada *pixel* de treinamento é atribuído a um neurônio de saída com base na máxima ativação obtida. Esse processo é repetido iterativamente e, em cada iteração, o resultado é comparado à solução prévia, sendo retida a estrutura de pesos que resultar no menor erro de teste. A iteração continua até que os pesos atinjam uma solução com erro de teste aceitável para particionar as variáveis de entrada em classes de saída, ou até que o usuário interrompa o processo de treinamento. Essa rede “treinada” é então usada para classificar o restante da área com base no nível de ativação do neurônio de saída induzido por um determinado *pixel* de entrada (Kanellopoulos & Wilkinson, 1997).

A Figura 26 mostra os parâmetros de treinamento da rede neural MLP para a densidade de amostragem de 1 ponto por hectare. Para as demais densidades o número mínimo de *pixels* de treinamento por classe foi estabelecido com base na distribuição do número de pontos por ordem de solo em cada densidade de amostragem (Tabela 5), adotando-se o menor valor verificado entre as quatro ordens. Em todas as densidades utilizou-se duas

camadas ocultas de 5 neurônios cada. O número de neurônios da camada de entrada e da camada de saída é atribuído automaticamente.

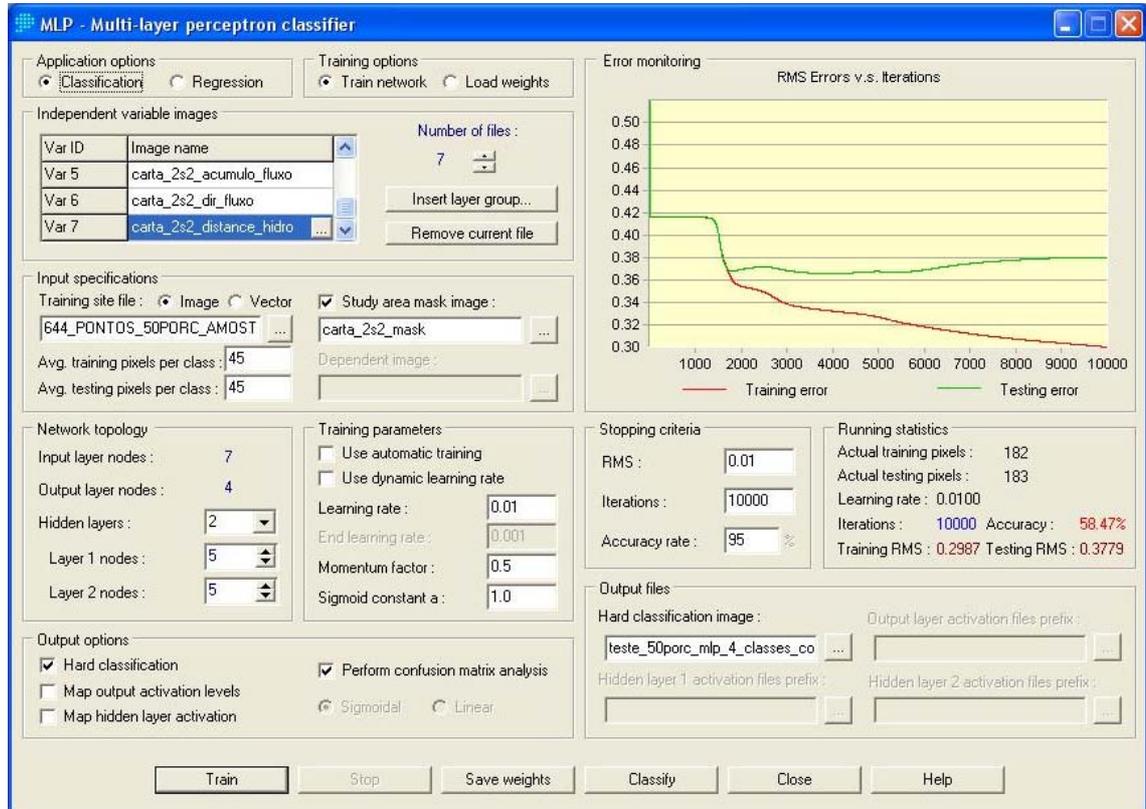


Figura 26. Parâmetros de treinamento da rede neural MLP com densidade de 1 ponto por hectare.

3.3.4.2. Rede neural *Adaptive Resonance Theory* (ART)

Redes neurais do tipo *Adaptive Resonance Theory* (ART) foram projetadas para modelar e resolver o dilema da plasticidade-estabilidade, simulando a habilidade do cérebro de aprender rapidamente e de forma estável sem esquecer o conhecimento previamente adquirido (Mas & Flores, 2008). Elas possuem um alto grau de estabilidade, possibilitando a retenção de grande parte do aprendizado passado, mas são suficientemente adaptáveis para incorporar nova informação na estrutura da rede sempre que for necessário.

Redes neurais ART incorporam um procedimento de aprendizado incremental, onde cada observação (*pixel*) da camada de entrada é atribuída a um neurônio da camada de saída. Se nenhum neurônio da camada de saída atender ao limiar de similaridade de uma dada observação da camada de entrada, um novo neurônio é criado na camada de saída a fim de particionar subconjuntos em um grau de homogeneidade definido pelo usuário através de

um parâmetro de “vigilância”. Redes neurais ART tem mostrado baixa sensibilidade aos parâmetros de treinamento e ao tamanho do conjunto de treinamento, *overfitting*, e também apresentam pequeno volume de intervenção necessário (Mannan et al., 1998; Chan et al., 2001).

A Figura 27 mostra os parâmetros de treinamento da rede neural ART para a densidade de amostragem de 1 ponto por hectare. Diferente da rede neural MLP, o número de neurônios de todas as camadas é atribuído automaticamente e varia com as diferentes densidades de amostragem. Da mesma forma, não é necessário definir um número mínimo de *pixels* de treinamento.

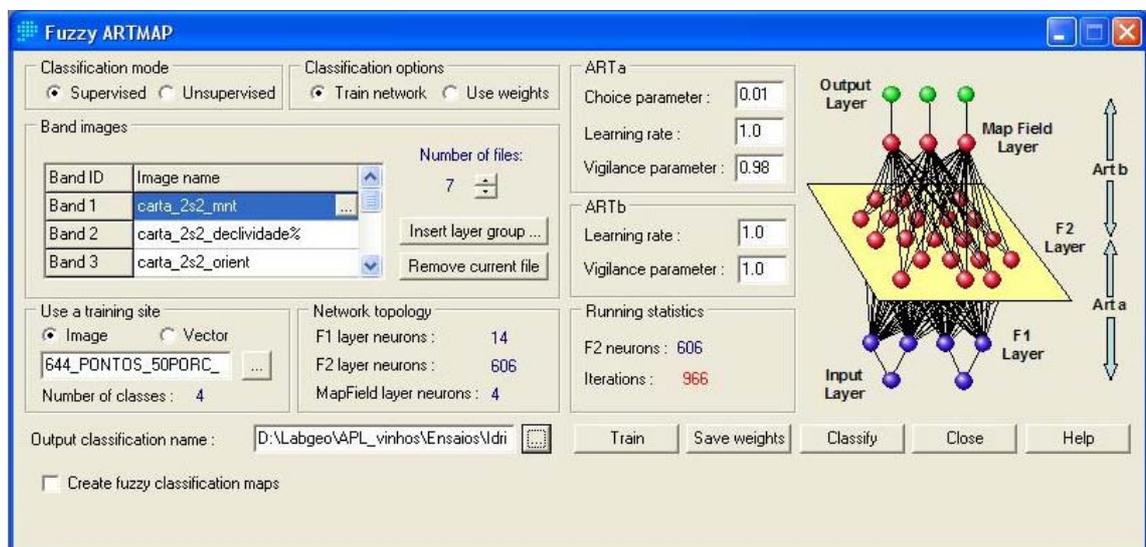


Figura 27. Parâmetros de treinamento da rede neural Fuzzy ARTMap com densidade de 1 ponto por hectare.

3.3.4.3. Rede neural *Self-Organizing Map* (SOM)

Redes neurais do tipo *Self-Organizing Map* (SOM) são um modelo de aspectos específicos das redes neurais biológicas, inspirado nas características dos mapas cognitivos no cérebro. Elas são baseadas em aprendizado competitivo, isto é, os neurônios competem entre si para serem ativados, e constituem um novo paradigma em inteligência artificial e modelagem cognitiva (Kohonen, 1990).

O processo de aprendizado de uma rede SOM ocorre em quatro fases: inicialização, competição, cooperação e adaptação sináptica. Nesse processo uma camada bidimensional única de neurônios é inicializada com pesos aleatórios e é subsequenteemente organizada através de uma amostragem sistemática dos dados de entrada. O procedimento de

organização usa ajustes progressivos dos pesos, com base nas características dos dados de entrada e no ajuste aos dados de saída, e possui interação lateral de forma que neurônios com pesos similares se agrupam espacialmente. O processo é repetido até que não sejam mais observadas mudanças, ou seja, até a convergência (Mas & Flores, 2008). Como padrões de entrada similares tendem a produzir resposta em neurônios vizinhos, redes SOM preservam as mais importantes relações topológicas e métricas dos dados. Assim, diferente de redes MLP e ART, redes SOM reconhecem relações entre classes, o que lhes confere uma alta capacidade para discriminação de classes multimodais e as torna um método promissor para a classificação de dados multivariados (Kohonen, 1990; Li & Eastman, 2006).

A Figura 28 mostra, como exemplo, os parâmetros de treinamento da rede neural SOM para a densidade de amostragem de 1 ponto por hectare. O número de neurônios na camada de entrada é atribuído automaticamente e não é necessário definir um número mínimo de *pixels* de treinamento. Em todas as densidades utilizou-se uma camada de saída com 45 neurônios.

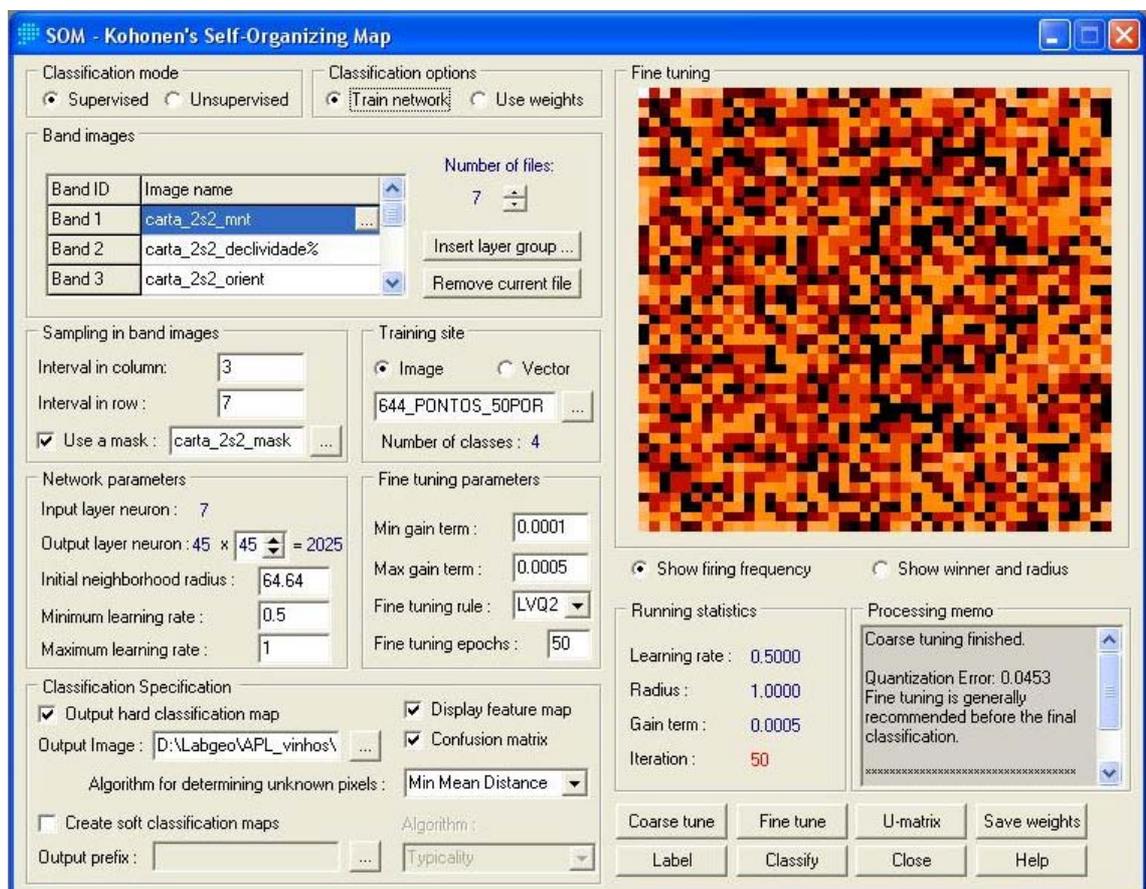


Figura 28. Parâmetros de treinamento da rede neural SOM com densidade de 1 ponto por hectare.

3.3.4.4. Árvore de decisão

Árvores de decisão ou de classificação usam um procedimento no qual os dados são particionados recursivamente para formar subconjuntos homogêneos, resultando em uma árvore hierárquica de regras de decisão. As árvores de decisão inicialmente analisam todas as variáveis de entrada e determinam uma divisão binária de uma variável isolada que minimize o desvio da variável dependente (classe) dos nós (bifurcações) recém-criados, ignorando particionamentos futuros e todos os dados de treinamento anteriormente particionados. Assim, enquanto o particionamento inicial é feito com todos os dados de entrada, cada partição subsequente é executada com um subconjunto de dados de treinamento cada vez menor, e por isso alheio às amostras fora do seu nó de origem. Da mesma forma que as redes neurais, árvores de decisão não assumem uma distribuição prévia dos dados e podem rapidamente caracterizar relações não lineares, com a vantagem de produzir regras de decisão facilmente interpretáveis (Rogan et al., 2003).

Existem vários algoritmos de particionamento nas árvores de decisão. Neste estudo foi utilizado o algoritmo Gini, que emprega uma medida de impureza para tentar encontrar a maior categoria homogênea dentro do conjunto de dados e isolá-la do restante dos dados, criando um novo nó (Zambon et. al., 2006). O índice Gini surgiu na economia, na tentativa de medir desigualdades. Aplicado à classificação de dados, uma categoria será tanto mais homogênea quanto menor for o índice Gini, que alcança o valor máximo quando todas as células estiverem igualmente distribuídas entre todas as classes.

A Figura 29 mostra, como exemplo, os parâmetros de treinamento da árvore de decisão para a densidade de amostragem de 1 ponto por hectare. Diferente das redes neurais, o único parâmetro de controle é a menor proporção ocupada por uma classe (folha), que foi fixada em 1% da área total.

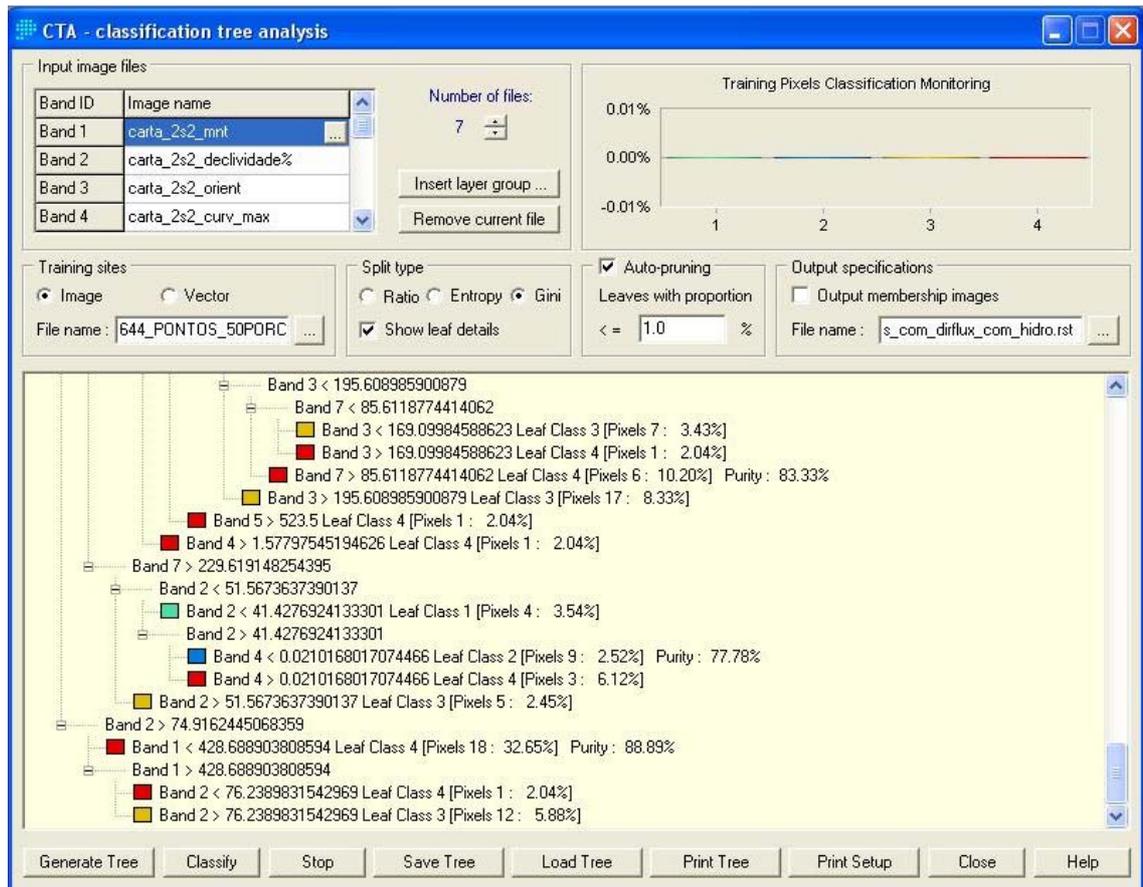


Figura 29. Parâmetros de treinamento da árvore de decisão Gini com densidade de 1 ponto por hectare.

3.3.5. Avaliação da acurácia

A concordância entre os mapas estimados com os quatro algoritmos e o mapa convencional de solos usado como referência foi avaliada através de matrizes de erro, da exatidão geral e do índice *Kappa* (Cohen, 1960; Pontius, 2000). A comparação entre cada mapa estimado e o mapa de referência foi efetuada com a rotina CROSSTAB, que executa uma análise de tabulação cruzada entre mapas matriciais contendo variáveis categóricas. O resultado é uma tabela completa de classificação cruzada, chamada de matriz de erro, acompanhada de uma estatística Chi quadrado e do respectivo nível de significância, além do índice *Kappa*.

4. RESULTADOS E DISCUSSÃO

4.1. Mapas estimados

A Figura 30 mostra o conjunto de 20 mapas estimados através da predição de ocorrência das ordens de solos com os quatro algoritmos de aprendizagem de máquina e as cinco densidades de amostragem. No topo da figura estão identificados os algoritmos e as letras na sua lateral esquerda estão relacionadas às densidades de amostragem correspondentes a cada mapa estimado.

Observando a Figura 30 constata-se que vários mapas estimados apresentam uma forte similaridade aparente, mas alguns aspectos, entretanto, podem ser destacados visualmente. A rede neural Fuzzy ARTMap, por exemplo, estimou a classe Argissolo em maior proporção que os outros três algoritmos em todas as densidades de amostragem inferiores a 4 pontos.ha⁻¹. A rede neural MLP, por outro lado, estimou apenas duas classes na densidade de amostragem de 0,5 ponto.ha⁻¹, Cambissolos e Chernossolos. Estas classes são as que apresentam maior ocorrência na área de estudo, respectivamente 41,77% e 33,96% da superfície, o que sugere que para este algoritmo a densidade de 0,5 ponto.ha⁻¹ gera um número de amostras de treinamento insuficiente para caracterizar as relações entre as variáveis preditoras e as ordens de solos com menor ocorrência.

Em todos os casos os mapas estimados apresentam maior detalhamento espacial que o mapa convencional de solos (Figura 17), ou seja, os algoritmos de classificação utilizados individualizaram delineamentos de solos menores que aquelas constantes no mapa convencional de solos. Esses resultados concordam com os estudos de Zhu (2000), Chagas et al. (2007), Hempel et al. (2008) e MacMillan (2008), onde os autores concluíram que os mapas estimados por métodos quantitativos tendem a mostrar muito mais detalhes espaciais do que o mapa convencional de solos. Entre as razões para essa diferença concorrem as limitações típicas do mapa convencional, como as generalizações decorrentes da necessidade de delimitar polígonos contendo unidades de mapeamento homogêneas e discretas, e a adoção de um limite de área mínima mapeável, fazendo com que unidades de solos com dimensões inferiores a esse limite sejam ignoradas ou incorporadas às classes predominantes no seu entorno.

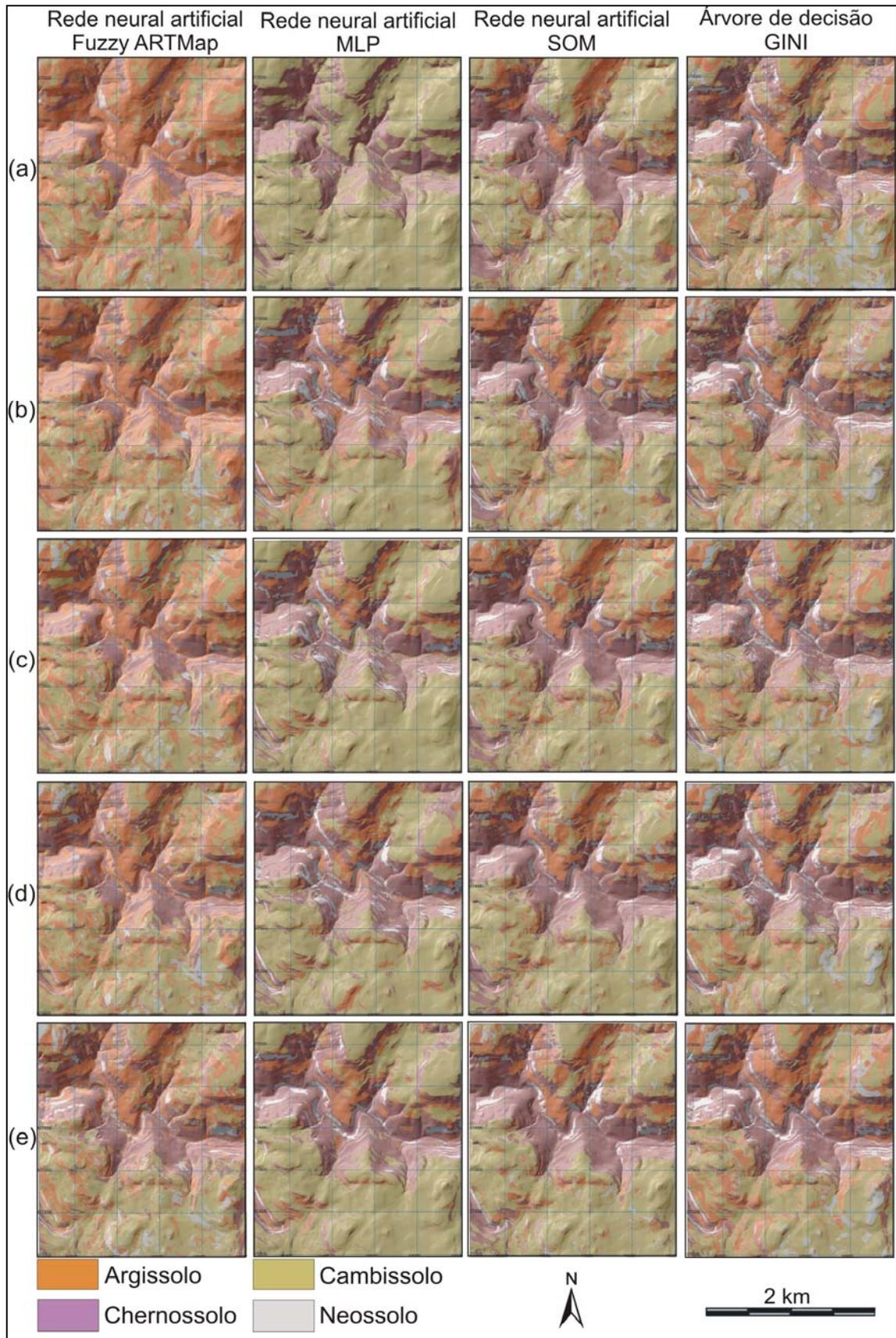


Figura 30. Mapas estimados através dos quatro algoritmos de aprendizagem de máquina e cinco densidades de amostragem: (a) 0,5 pontos.ha⁻¹; (b) 1 ponto.ha⁻¹; (c) 1,5 pontos.ha⁻¹; (d) 2 pontos.ha⁻¹; (e) 4 pontos.ha⁻¹.

Em contrapartida, a análise das variáveis preditoras em SIG para estimar os solos é executada *pixel a pixel*, de forma que as dimensões deste, e não de um polígono, constituem a menor unidade de mapeamento. Outro aspecto relevante é que, uma vez estabelecido um modelo quantitativo para predição dos solos a partir das variáveis selecionadas, ele é aplicado uniformemente sobre toda a área a ser mapeada. No mapeamento convencional freqüentemente há locais inacessíveis ou que não foram visitados, para os quais torna-se necessário inferir as relações solo-paisagem a partir de observações feitas em outros locais, induzindo a uma maior generalização na delimitação dos polígonos de solos.

Assim, é razoável assumir que as variáveis preditoras utilizadas permitiram estabelecer com maior precisão as relações solo-paisagem que ocorrem na área de estudo, possibilitando a obtenção de mapas espacialmente mais detalhados. O maior detalhamento é um aspecto positivo dos mapas estimados, uma vez que permite reconhecer unidades de mapeamento impossíveis ou difíceis de delimitar em um mapa convencional.

4.2. Matrizes de erro

As tabelas 6 a 25 mostram as matrizes de erro obtidas através da comparação de cada mapa estimado com o mapa convencional de solos. Em todos os casos, Chi quadrado foi significativo em nível de probabilidade inferior a 1% ($p < 0,01$). Entretanto, como a rede neural MLP estimou apenas as duas classes de maior ocorrência na densidade de amostragem de $0,5 \text{ ponto.ha}^{-1}$, a comparação dessa matriz de erro (Tabela 7) com as demais deve ser feita com cautela.

Os valores nas matrizes de erro estão expressos como proporção da área total mapeada, onde as colunas mostram os dados da verdade de campo (classes do mapa convencional de solos) e as linhas o resultado da predição (classes estimadas). O total da coluna mostra a proporção da área ocupada pela respectiva classe no mapa convencional de solos e o total da linha mostra a proporção da respectiva classe no mapa estimado.

Tabela 6. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 0,5 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0971	0,1807	0,1978	0,0586	0,5342	0,8182
	Cambissolo	0,0446	0,2035	0,0243	0,0173	0,2897	0,2975
	Chernossolo	0,0087	0,0276	0,1107	0,0113	0,1583	0,3007
	Neossolo	0,0004	0,0059	0,0068	0,0047	0,0179	0,7318
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,3561	0,5128	0,6740	0,9489		

Tabela 7. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 0,5 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	--	--	--	--	--	--
	Cambissolo	0,1364	0,4022	0,1071	0,0434	0,6892	0,4163
	Chernossolo	0,0144	0,0155	0,2325	0,0485	0,3108	0,3372
	Neossolo	--	--	--	--	--	--
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		1,0000	0,0371	0,3154	1,0000		

Tabela 8. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 0,5 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0415	0,0452	0,0386	0,0061	0,1313	0,6847
	Cambissolo	0,0813	0,2991	0,0627	0,0332	0,4764	0,3719
	Chernossolo	0,0236	0,0534	0,2280	0,0366	0,3416	0,3325
	Neossolo	0,0044	0,0200	0,0102	0,0160	0,0506	0,6838
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,7248	0,2839	0,3283	0,8259		

Tabela 9. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 0,5 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0694	0,0845	0,0365	0,0094	0,1999	0,6523
	Cambissolo	0,0541	0,2689	0,0333	0,0188	0,3752	0,2830
	Chernossolo	0,0143	0,0324	0,2444	0,0234	0,3145	0,2229
	Neossolo	0,0129	0,0318	0,0254	0,0403	0,1104	0,6350
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,5391	0,3560	0,2803	0,5615		

Tabela 10. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 1 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0927	0,1545	0,1773	0,0531	0,4776	0,8021
	Cambissolo	0,0401	0,2161	0,0287	0,0175	0,3025	0,2853
	Chernossolo	0,0162	0,0243	0,1271	0,0117	0,1793	0,2911
	Neossolo	0,0018	0,0227	0,0065	0,0095	0,0405	0,7654
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,3853	0,4824	0,6257	0,8955		

Tabela 11. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 1 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0500	0,0532	0,0493	0,0031	0,1556	0,6787
	Cambissolo	0,0844	0,3370	0,0255	0,0365	0,4833	0,3029
	Chernossolo	0,0156	0,0270	0,2375	0,0249	0,3050	0,2213
	Neossolo	0,0008	0,0005	0,0273	0,0275	0,0561	0,5098
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,6684	0,1932	0,3006	0,7018		

Tabela 12. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 1 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0625	0,0732	0,0531	0,0091	0,1979	0,6842
	Cambissolo	0,0633	0,2932	0,0448	0,0291	0,4304	0,3188
	Chernossolo	0,0233	0,0381	0,2239	0,0346	0,3199	0,3001
	Neossolo	0,0017	0,0131	0,0178	0,0192	0,0518	0,6293
	Total	0,1508	0,4177	0,3396	0,0919		
	Omissão	0,5855	0,2978	0,3407	0,7922		

Tabela 13. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 1 ponto.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0820	0,0617	0,0322	0,0095	0,1854	0,5577
	Cambissolo	0,0423	0,3131	0,0435	0,0153	0,4142	0,2441
	Chernossolo	0,0178	0,0203	0,2301	0,0253	0,2935	0,2160
	Neossolo	0,0087	0,0226	0,0338	0,0417	0,1069	0,6070
	Total	0,1508	0,4177	0,3396	0,0919		
	Omissão	0,4562	0,2504	0,3224	0,5452		

Tabela 14. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 1,5 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0934	0,1347	0,1201	0,0404	0,3886	0,7595
	Cambissolo	0,0362	0,2407	0,0317	0,0209	0,3295	0,2695
	Chernossolo	0,0154	0,0237	0,1796	0,0125	0,2312	0,2232
	Neossolo	0,0058	0,0186	0,0082	0,0181	0,0507	0,6430
	Total	0,1508	0,4177	0,3396	0,0919		
	Omissão	0,3806	0,4237	0,4711	0,8030		

Tabela 15. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 1,5 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0367	0,0203	0,0194	0,0030	0,0793	0,5385
	Cambissolo	0,0996	0,3791	0,0628	0,0388	0,5803	0,3467
	Chernossolo	0,0145	0,0181	0,2368	0,0262	0,2956	0,1989
	Neossolo	0,0001	0,0002	0,0206	0,0239	0,0448	0,4665
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,7573	0,0924	0,3027	0,7399		

Tabela 16. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 1,5 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0705	0,0631	0,0439	0,0086	0,1861	0,6212
	Cambissolo	0,0555	0,3074	0,0438	0,0308	0,4375	0,2974
	Chernossolo	0,0227	0,0368	0,2405	0,0327	0,3326	0,2772
	Neossolo	0,0022	0,0104	0,0114	0,0197	0,0438	0,5479
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,5331	0,2641	0,2918	0,7845		

Tabela 17. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 1,5 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0874	0,0747	0,0292	0,0096	0,2009	0,5650
	Cambissolo	0,0375	0,2884	0,0276	0,0143	0,3678	0,2159
	Chernossolo	0,0205	0,0278	0,2521	0,0264	0,3268	0,2286
	Neossolo	0,0054	0,0267	0,0307	0,0416	0,1044	0,6015
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,4204	0,3093	0,2577	0,5473		

Tabela 18. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 2 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0936	0,1085	0,1010	0,0363	0,3394	0,7242
	Cambissolo	0,0372	0,2644	0,0365	0,0191	0,3572	0,2597
	Chernossolo	0,0170	0,0269	0,1914	0,0147	0,2499	0,2341
	Neossolo	0,0031	0,0178	0,0107	0,0218	0,0535	0,5931
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,3799	0,3668	0,4364	0,7632		

Tabela 19. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 2 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0533	0,0472	0,0184	0,0063	0,1252	0,5743
	Cambissolo	0,0711	0,3272	0,0318	0,0327	0,4629	0,2929
	Chernossolo	0,0257	0,0430	0,2654	0,0254	0,3595	0,2617
	Neossolo	0,0007	0,0002	0,0240	0,0274	0,0524	0,4752
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,6465	0,2164	0,2185	0,7008		

Tabela 20. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 2 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0657	0,0448	0,0336	0,0067	0,1507	0,5647
	Cambissolo	0,0620	0,3267	0,0464	0,0308	0,4659	0,2988
	Chernossolo	0,0204	0,0365	0,2497	0,0323	0,3389	0,2632
	Neossolo	0,0027	0,0097	0,0099	0,0221	0,0445	0,5011
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,5643	0,2179	0,2647	0,7595		

Tabela 21. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 2 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0921	0,0521	0,0182	0,0074	0,1698	0,4578
	Cambissolo	0,0387	0,3200	0,0351	0,0144	0,4083	0,2163
	Chernossolo	0,0141	0,0256	0,2561	0,0252	0,3211	0,2024
	Neossolo	0,0059	0,0200	0,0301	0,0448	0,1008	0,5558
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,3894	0,2339	0,2459	0,5129		

Tabela 22. Matriz de erro obtida com o algoritmo de rede neural Fuzzy ARTMap e densidade de amostragem de 4 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0971	0,0703	0,0659	0,0263	0,2596	0,6260
	Cambissolo	0,0363	0,2992	0,0307	0,0199	0,3861	0,2251
	Chernossolo	0,0132	0,0305	0,2311	0,0158	0,2906	0,2047
	Neossolo	0,0042	0,0177	0,0120	0,0299	0,0637	0,5322
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,3561	0,2837	0,3198	0,6746		

Tabela 23. Matriz de erro obtida com o algoritmo de rede neural MLP e densidade de amostragem de 4 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0431	0,0224	0,0203	0,0009	0,0867	0,5029
	Cambissolo	0,0917	0,3734	0,0401	0,0392	0,5444	0,3141
	Chernossolo	0,0149	0,0217	0,2686	0,0275	0,3327	0,1927
	Neossolo	0,0012	0,0002	0,0106	0,0243	0,0362	0,3315
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,7148	0,1061	0,2091	0,7356		

Tabela 24. Matriz de erro obtida com o algoritmo de rede neural SOM e densidade de amostragem de 4 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,0653	0,0403	0,0326	0,0049	0,1431	0,5437
	Cambissolo	0,0619	0,3322	0,0394	0,0315	0,4650	0,2856
	Chernossolo	0,0199	0,0337	0,2552	0,0312	0,3400	0,2494
	Neossolo	0,0037	0,0114	0,0124	0,0244	0,0518	0,5309
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,5670	0,2044	0,2485	0,7356		

Tabela 25. Matriz de erro obtida com o algoritmo de Árvore de decisão Gini e densidade de amostragem de 4 pontos.ha⁻¹.

		Referência					
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	Inclusão
Estimado	Argissolo	0,1114	0,0866	0,0500	0,0100	0,2580	0,5682
	Cambissolo	0,0226	0,2843	0,0202	0,0132	0,3402	0,1646
	Chernossolo	0,0081	0,0251	0,2453	0,0232	0,3017	0,1869
	Neossolo	0,0088	0,0217	0,0241	0,0455	0,1001	0,5454
	Total	0,1508	0,4177	0,3396	0,0919		
Omissão		0,2619	0,3194	0,2777	0,5049		

Os dados das tabelas 6 a 25 mostram que o algoritmo de árvore de decisão Gini foi o que melhor conseguiu reproduzir no mapa estimado (total das linhas) a distribuição espacial e a proporção ocupada pelas diferentes classes no mapa convencional de solos (total das colunas), com a densidade de amostragem de 2 pontos.ha⁻¹ (Tabela 21). Neste caso a classe Argissolo totaliza 16,98% da área de estudo no mapa estimado e 15,08% no mapa convencional, a classe Cambissolo 40,83% no mapa estimado e 41,77% no convencional, a classe Chernossolo 32,11% no mapa estimado e 33,96% no convencional e a classe Neossolo 10,8% no mapa estimado e 9,19% no mapa convencional.

Os erros de inclusão (linhas) variaram de 0,1646, para a classe Cambissolos estimada com o algoritmo de árvore de decisão Gini na densidade de amostragem de 4 pontos.ha⁻¹ (Tabela 25), a 0,8182, para a classe Argissolos estimada com o algoritmo de rede neural Fuzzy ARTMap na

densidade de amostragem de 0,5 ponto.ha⁻¹ (Tabela 6). Os erros de omissão (colunas) variaram de 0,0924, para a classe Cambissolos estimada com o algoritmo de rede neural MLP na densidade de amostragem de 1,5 pontos.ha⁻¹ (Tabela 15), a 0,9489, para a classe Neossolos estimada com o algoritmo de rede neural Fuzzy ARTMap na densidade de 0,5 ponto.ha⁻¹ (Tabela 6).

Considerando-se as classes estimadas, os menores erros de inclusão para as classes Argissolo (Tabela 21), Cambissolo (Tabela 25) e Chernossolo (Tabela 25) foram obtidos com a árvore de decisão Gini, respectivamente nas densidades de amostragem de 2, 4 e 4 pontos.ha⁻¹, enquanto o menor erro de inclusão para a classe Neossolo (Tabela 23) foi obtido com a rede neural MLP na densidade de 4 pontos.ha⁻¹. Com relação aos erros de omissão, os menores valores para as classes Argissolo e Neossolo foram obtidos com a árvore de decisão Gini (Tabela 25) na densidade de amostragem de 4 pontos.ha⁻¹ e para as classes Cambissolo (Tabela 15) e Chernossolo (Tabela 23) pela rede neural MLP, respectivamente nas densidades de 1,5 e 4 pontos.ha⁻¹.

As classes Cambissolo e Chernossolo apresentaram os menores erros de omissão e de inclusão em praticamente todas as densidades de amostragem e algoritmos. Esse resultado é consistente com o esperado já que as duas classes ocupam a maior superfície, respectivamente 41,77% e 33,96% da área de estudo, e têm, portanto, maior probabilidade de serem corretamente mapeadas. A exceção ocorreu com a rede neural Fuzzy ARTMap, onde os menores erros de omissão foram obtidos para as classes Argissolo e Cambissolo em todas as densidades de amostragem inferiores a 4 pontos.ha⁻¹. Entretanto, em oposição aos baixos erros de omissão, os erros de inclusão para os Argissolos neste caso foram mais altos que em todos os demais algoritmos, sugerindo uma superestimativa desta classe. Como os Argissolos ocupam apenas 15,08% da área de estudo, é possível que as densidades de amostragem inferiores a 4 pontos.ha⁻¹ não tenham sido suficientes para caracterizar satisfatoriamente as relações solo-paisagem desta classe para a rede neural Fuzzy ARTMap.

Em todos os quatro algoritmos testados observa-se que, à medida que aumenta a densidade de amostragem, tanto os erros de inclusão quanto os erros de omissão tendem a diminuir, para todas as classes estimadas. Isso

significa que à medida que são fornecidas mais amostras de treinamento para os algoritmos, as relações entre as variáveis preditoras e as classes em questão são mais bem caracterizadas, tornando as regras de classificação mais consistentes e diminuindo a confusão entre as classes.

4.3. Exatidão geral

A Tabela 26 relaciona a exatidão geral obtida com os quatro algoritmos de aprendizagem de máquina e as cinco densidades de amostragem. Como as matrizes de erro já estão expressas em proporção, a exatidão geral é obtida pela soma simples dos elementos da diagonal principal de cada matriz. Os valores variaram de 42%, com a rede Neural Fuzzy ARTMap e densidade de amostragem de 0,5 ponto.ha⁻¹, a 71%, exatidão obtida tanto com a árvore de decisão Gini, na densidade de amostragem de 2 pontos.ha⁻¹, quanto com a Rede neural MLP, na densidade de amostragem de 4 pontos.ha⁻¹. Com a densidade de amostragem mais elevada (4 pontos.ha⁻¹) os quatro algoritmos convergiram (Figura 31) e apresentaram valores próximos, variando de 66 a 71%.

Tabela 26. Exatidão geral (%) dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.

Algoritmo de classificação	Densidade de amostragem (pontos.ha ⁻¹)					Média
	0,5	1	1,5	2	4	
Rede neural Fuzzy ARTMap	42	45	53	57	66	52,6
Rede neural MLP	----*	65	68	67	71	67,8
Rede neural SOM	58	60	64	66	68	63,2
Árvore de decisão (Gini)	62	67	67	71	69	67,2
Média	54,0	59,3	63,0	65,3	68,5	

* Estimou apenas duas classes: Cambissolos e Chernossolos

A Figura 31 mostra que a árvore de decisão Gini e a rede neural MLP apresentaram exatidão geral semelhante em todas as densidades de amostragem, com valores superiores aos das redes neurais Fuzzy ARTMap e SOM. Além disso, os dois algoritmos apresentam uma curva com pouca inclinação, denotando uma resposta pouco acentuada ao incremento na densidade de amostragem. Esses resultados sugerem que a árvore de decisão

Gini e a rede neural MLP são menos sensíveis à densidade de amostragem que as outras duas redes neurais, embora a rede neural MLP não tenha sido capaz de estimar todas as classes na densidade de amostragem de $0,5 \text{ pontos.ha}^{-1}$. Esse resultado sugere que a rede neural MLP tem um limite inferior crítico, enquanto os demais algoritmos apenas reduzem a acurácia à medida que diminui a densidade de amostragem. A rede neural Fuzzy ARTMap foi o algoritmo com a resposta mais acentuada ao incremento na densidade de amostragem, como pode ser constatado pela curva acentuadamente inclinada.

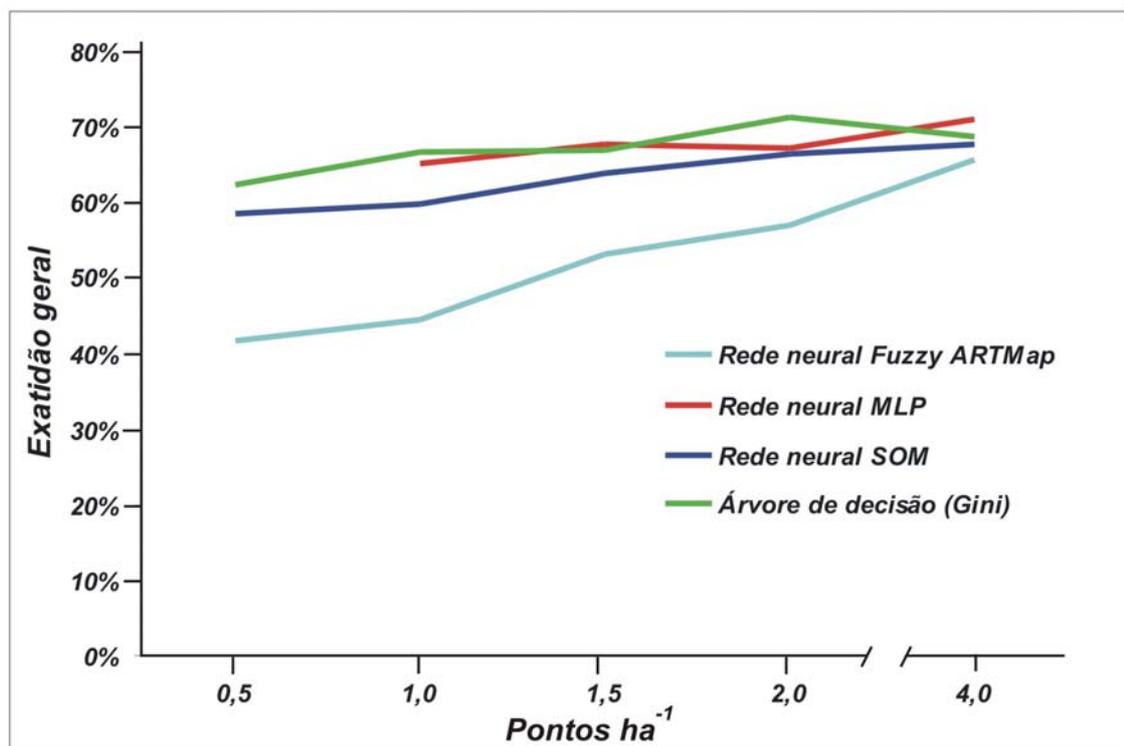


Figura 31. Exatidão geral dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.

4.4. Índice *Kappa*

A Tabela 27 relaciona o índice *Kappa* obtido com os quatro algoritmos de aprendizagem de máquina e as cinco densidades de amostragem. O índice *Kappa* expressa o grau de concordância entre os mapas estimados e o mapa de referência, após a concordância devido ao acaso ter sido eliminada.

Tabela 27. Índice *Kappa* dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.

Algoritmo de classificação	Densidade de amostragem (pontos.ha ⁻¹)					Média
	0,5	1	1,5	2	4	
Rede neural Fuzzy ARTMap	0,214	0,2476	0,3502	0,3959	0,5069	0,3429
Rede neural MLP	----*	0,4775	0,4955	0,5057	0,5482	0,5068
Rede neural SOM	0,3711	0,4073	0,4616	0,4940	0,5135	0,4495
Árvore de decisão (Gini)	0,4586	0,5169	0,5248	0,5813	0,5566	0,5276
Média	0,3479	0,4124	0,4581	0,4943	0,5313	

* Estimou apenas duas classes: Cambissolos e Chernossolos

O índice *Kappa* variou de 0,214, para a rede neural Fuzzy ARTMap na densidade de amostragem de 0,5 ponto.ha⁻¹, a 0,5813, para a árvore de decisão Gini na densidade de 2 pontos.ha⁻¹. Entre os algoritmos de redes neurais, a melhor concordância foi obtida com a rede neural MLP na densidade de amostragem de 4 pontos.ha⁻¹, com um valor de *Kappa* de 0,5482, seguida da rede neural SOM, com *Kappa* de 0,5135, também na densidade de 4 pontos.ha⁻¹. Com a densidade de amostragem mais elevada (4 pontos.ha⁻¹) os quatro algoritmos convergiram (Figura 32) e apresentaram valores próximos, variando de 0,5069 a 0,5566.

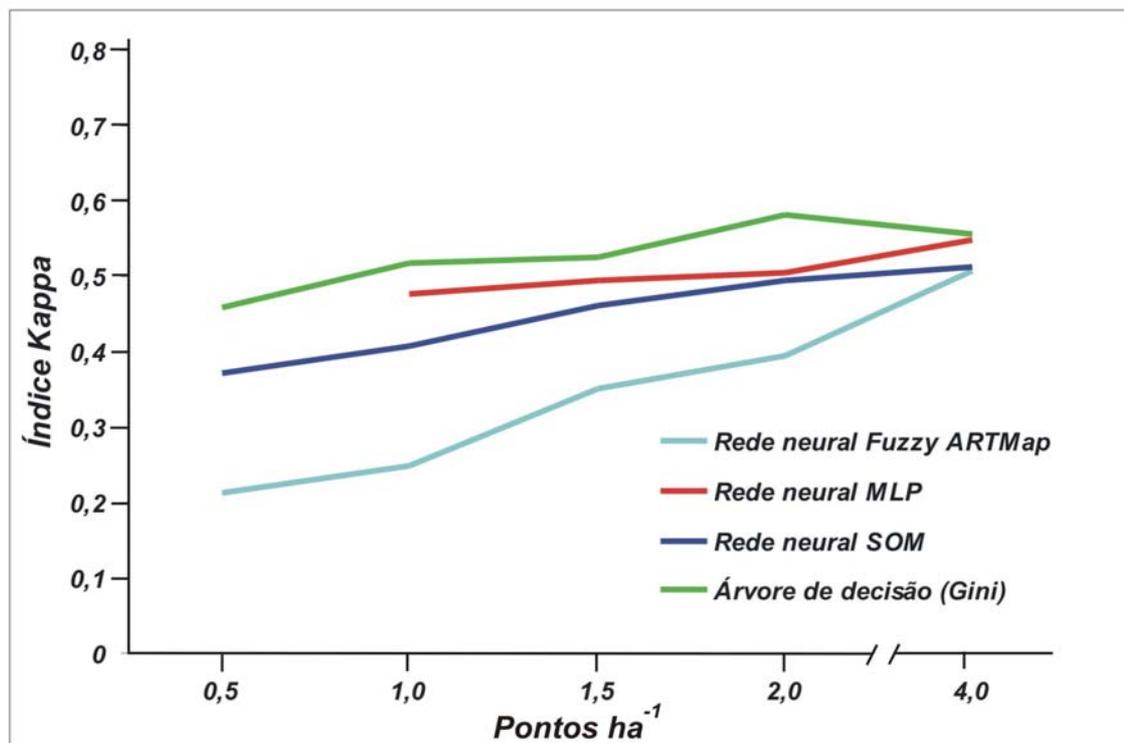


Figura 32. Índice *Kappa* dos mapas estimados através dos quatro algoritmos de aprendizagem de máquina em cinco densidades de amostragem.

Segundo a escala sugerida por Luoto & Hjort (2005), nenhum dos mapas estimados apresentou concordância forte (0,61 – 0,80) com o mapa convencional usado como referência, mas a árvore de decisão Gini chegou próximo do limite, com *Kappa* de 0,5813, e apresentou concordância moderada em todas as densidades de amostragem. O mapa estimado com a rede neural MLP apresentou concordância moderada com densidade de amostragem a partir de 1 ponto.ha⁻¹, a rede neural SOM a partir de 1,5 pontos.ha⁻¹ e a rede neural Fuzzy ARTMap apenas com 4 pontos.ha⁻¹.

Os dados da Tabela 27 e da Figura 32 mostram também que a árvore de decisão Gini apresentou índice *Kappa* superior aos três algoritmos de redes neurais em todas as densidades de amostragem. Em outras palavras, retirado o acaso, a árvore de decisão Gini obteve concordância superior aos três algoritmos de redes neurais, o que não é evidente na exatidão geral, que desconsidera o efeito do acaso. Além disso, a árvore de decisão apresentou índice *Kappa* acima do limiar de 0,5, considerado bom, em todas as densidades de amostragem a partir de 1 ponto.ha⁻¹. Entre os três algoritmos de redes neurais, a rede neural MLP apresentou índice *Kappa* acima de 0,5 com densidade de amostragem a partir de 2 pontos.ha⁻¹, embora com 1,5 ponto.ha⁻¹ praticamente tenha atingido esse limiar. As redes neurais Fuzzy ARTMap e SOM apresentaram índice *Kappa* acima de 0,5 somente com densidade de amostragem de 4 pontos.ha⁻¹.

Quanto à sensibilidade à densidade de amostragem, a resposta dos algoritmos medida pelo índice *Kappa* praticamente não se alterou em relação aos resultados expressos na exatidão geral. A árvore de decisão Gini e a rede neural MLP foram menos sensíveis à densidade de amostragem, enquanto a rede neural Fuzzy ARTMap apresentou a resposta mais acentuada ao incremento na densidade de amostragem. Entretanto, a árvore de decisão Gini apresentou exatidão geral e índice *Kappa* decrescentes após 2 pontos.ha⁻¹, sugerindo que o limite superior para o número de amostras de treinamento foi atingido. É necessário testar densidades de amostragem maiores para verificar se isto é verdadeiro.

Na etapa final da avaliação foi realizada a comparação entre o desempenho da classificação pelos dois algoritmos menos sensíveis à densidade de amostragem. A Tabela 28 mostra a matriz de erro entre o mapa

estimado com a árvore de decisão Gini com densidade de amostragem de 2 pontos.ha⁻¹ (índice *Kappa* de 0,5813) e o mapa estimado com a rede neural MLP com densidade de amostragem de 4 pontos⁻¹ (índice *Kappa* de 0,5482). Chi quadrado foi significativo para nível de probabilidade inferior a 1% (p<0,01).

Considerando-se a árvore de decisão Gini como referência, a exatidão geral do mapa estimado com a rede neural MLP foi de 73,22%, ou seja, da superfície total de 673,5 ha da área de estudo, 493,14 ha foram coincidentes entre os mapas estimados com os dois algoritmos. O índice *Kappa* foi de 0,5896, o que demonstra que a concordância entre os mapas obtidos com os dois algoritmos é maior do que aquela obtida devido ao acaso. Embora diferentes quanto à natureza dos seus modelos matemáticos, os dois algoritmos mostraram-se capazes de mapear de forma semelhante uma parcela expressiva das quatro classes na área de estudo. A classe Chernossolos foi a mais coincidente entre os dois algoritmos, totalizando 32,11% da área de estudo no mapa estimado com a árvore de decisão Gini e 33,27% no mapa estimado com a rede neural MLP.

Tabela 28. Matriz de erro entre o mapa estimado com a árvore de decisão Gini com densidade de amostragem de 2 pontos.ha⁻¹ (colunas) e o mapa estimado com a rede neural MLP com densidade de amostragem de 4 pontos.ha⁻¹ (linhas).

		Árvore de decisão Gini					Inclusão
		Argissolo	Cambissolo	Chernossolo	Neossolo	Total	
Rede neural MLP	Argissolo	0,0484	0,0197	0,0186	0,0000	0,0867	0,4418
	Cambissolo	0,1064	0,3800	0,0163	0,0417	0,5444	0,3020
	Chernossolo	0,0151	0,0086	0,2769	0,0321	0,3327	0,1677
	Neossolo	0,0000	0,0000	0,0093	0,0269	0,0362	0,2569
	Total	0,1698	0,4083	0,3211	0,1008		
Omissão		0,7155	0,0691	0,1377	0,7321		

4.5. Discussão geral

4.5.1. Sobre as variáveis preditoras

A escolha das variáveis preditoras é uma tarefa complexa e requer um bom grau de conhecimento sobre a importância relativa dos fatores de formação de solos na região de estudo (Zhu, 2000), além de depender de informações disponíveis. O conjunto de variáveis utilizado neste estudo não

pode ser considerado completo em termos de fatores de formação dos solos, mas procurou incluir aquelas que caracterizem o ambiente de formação dos solos em áreas declivosas e que sejam possíveis de derivar e analisar em ambiente de Sistema de Informação Geográfica.

Variáveis climáticas não foram utilizadas porque o clima dominante é homogêneo para a escala do estudo, embora o relevo local propicie variações locais no clima do solo, como alteração da temperatura e umidade do solo. Em função do relevo movimentado, as variações microclimáticas são pronunciadas, mas são bem expressas pelas variáveis do terreno, principalmente pela elevação e orientação (Zhu, 2000). A vegetação, que poderia representar os organismos, não foi utilizada porque a formação original dominante, a Floresta Estacional Decidual (IBGE, 1986), é homogênea para a área de estudo, além de encontrar-se muito alterada pela ação antrópica. Esta por sua vez, é altamente condicionada pelo relevo, especialmente pela declividade. O material de origem, representado pela geologia, também não foi considerado por ser homogêneo para as escalas de mapeamento disponíveis (IBGE, 1986). Não há mapas em escala detalhada que possam representar eventuais variações dentro da área de estudo. Por último, o fator tempo não foi contemplado porque não há uma forma definida de computar esse fator para cada *pixel* da área de estudo, embora Zhu (2000) e Chagas et al. (2007) lembrem que muitas vezes ele esteja implícito em outros fatores.

A avaliação de acurácia mostrou que o conjunto de variáveis utilizado foi adequado, uma vez que a exatidão geral e o índice *Kappa* encontrados situam-se em patamares considerados satisfatórios. Como o relevo é o fator preponderante na área de estudo, variáveis que representam bem a forma e energia do relevo tendem a funcionar como boas preditoras dos tipos de solos. Conforme destacado por Rossiter (2004a), o fato de aproveitar o conhecimento de especialistas com experiência local (Flores et al., 2007; Sarmiento et al., 2008) na seleção de variáveis que caracterizam as forças dominantes de formação dos solos também colabora neste sentido.

Provavelmente a inclusão de variáveis adicionais possa elevar a acurácia dos mapas estimados, uma vez que as variáveis preditoras utilizadas não contemplam todos os fatores de formação e não explicam totalmente a distribuição das ordens de solos na área de estudo. Entretanto, identificar

variáveis complementares relevantes, evitando a inclusão de dados pouco significativos ou redundantes, requer uma avaliação prévia com técnicas apropriadas. Evitar a sobreposição de dados é importante para prevenir uma demanda de processamento desnecessário e uma caracterização deficiente das relações entre as variáveis selecionadas e as classes desejadas. Nesse sentido, a adoção de uma etapa de pré-processamento tem sido sugerida como alternativa para reduzir a dimensionalidade dos dados e elevar a qualidade das predições (Qi & Zhu, 2003; Qi, 2004; Ehsani & Quiel, 2008; Kavzoglu, 2009, Lopez & Demattê, 2009).

4.5.2. Sobre os mapas estimados e o mapa convencional de solos

É importante ressaltar que os resultados encontrados expressam a capacidade dos algoritmos testados em reproduzir o mapa convencional de solos usado como referência. Em função das incertezas inerentes à delimitação das unidades de solos no mapa convencional, é provável que se obtenha acurácia mais elevada caso sejam utilizadas observações de referência feitas em campo para treinar os algoritmos e avaliar os mapas estimados, em vez de se extrair os dados de polígonos de um mapa convencional existente.

Segundo Lopez (2009), os erros inerentes aos mapas convencionais interferem no treinamento, reduzindo o desempenho de predição. De acordo com Zhu (2000), a generalização típica do mapa convencional resulta em incompatibilidades entre a classe de solo e os atributos do terreno em muitos locais, o que introduz inconsistências nas amostras situadas próximo aos limites entre duas classes distintas. Assim, quanto mais próximo ao limite entre duas classes, maior tende a ser a chance de discordância entre a classe estimada e a classe do mapa convencional (Grimm & Behrens, 2010). Conforme destacado por Zhou et al. (2004), o ideal é realizar o teste de acurácia com pontos observados em campo, mas restrições de tempo, custo e pessoal freqüentemente tornam inviável a coleta de pontos em número suficiente para essa finalidade.

Nesse contexto, como o objetivo do estudo foi avaliar e comparar o desempenho dos algoritmos, o uso do mapa convencional como referência justifica-se pelo custo e tempo que seriam necessários para levantar informações em campo em todos os pontos de amostragem, o que acabaria

por inviabilizar o estudo. Os resultados encontrados fornecem, ainda assim, indicativos valiosos para racionalizar a coleta de dados em estudos futuros com observações em campo.

4.5.3. Sobre as densidades de amostragem

A árvore de decisão Gini foi o algoritmo que apresentou a menor sensibilidade à densidade de amostragem, resultado que concorda com trabalhos anteriores sugerindo que árvores de decisão são capazes de obter boa acurácia mesmo com conjuntos reduzidos de amostras de treinamento (Pal & Mather, 2002; Lippitt et al., 2008). A maior sensibilidade verificada para as redes neurais concorda com o observado por Richards & Jia (1999) de que o número de amostras de treinamento afeta a acurácia do resultado das redes neurais. Assim, do ponto de vista prático, a árvore de decisão Gini mostra-se mais vantajosa para a predição de ordens de solos na área de estudo em escala detalhada, pois permite obter resultados de boa qualidade com menor densidade de amostragem que as redes neurais.

Levar em conta o tamanho do conjunto amostral torna-se crítico, não apenas pelo esforço que a coleta de amostras representa, mas porque a sensibilidade de diferentes métodos de predição tende a se acentuar quando os dados apresentam relações não lineares, o que frequentemente ocorre no mapeamento digital de solos. Segundo Kavzoglu (2009), no caso de algoritmos de aprendizagem de máquina, uma classificação bem sucedida depende do número de variáveis em questão e do tipo de dados usado para representar cada uma, além do número e da representatividade das amostras de treinamento, entre outros fatores.

Zhu (2000) sugere adotar como limite inferior um número de amostras equivalente a pelo menos trinta vezes o número de classes que se deseja estimar, independente da escala. Hjort (2008), por outro lado, afirma que o tamanho da amostra afeta significativamente a capacidade de predição dos algoritmos, cuja acurácia tende a aumentar com a adição de mais informação, até atingir um platô. A partir desse ponto a adição de novos dados resulta em um incremento cada vez menor na acurácia, podendo inclusive ocorrer uma inflexão caso o aumento indiscriminado no número de observações promova a redundância e a sobreposição dos dados. Portanto, obter informações que sirvam de referência e auxiliem na tomada de decisão

sobre o volume de dados de entrada necessários para a obtenção de um resultado aceitável com um determinado algoritmo é extremamente útil.

Neste sentido, Perlich et al. (2003) conduziram uma extensa e aprofundada comparação entre árvores de decisão e regressões logísticas, avaliando suas curvas de aprendizado. Considerando um mesmo domínio, os autores constataram que regressões logísticas funcionam melhor com conjuntos de dados menores, enquanto árvores de decisão suportam extensos volumes de dados, mas há um ponto em que suas curvas de aprendizado se cruzam. Os autores concluem que sem examinar as curvas de aprendizado, afirmar que um determinado método é superior a outro em um domínio específico é, no mínimo, questionável.

Os resultados encontrados mostraram que os quatro algoritmos testados tendem a convergir a partir da densidade de amostragem de 4 pontos.ha⁻¹, o que sugere que as densidades recomendadas para levantamentos convencionais também podem ser adotadas como referência para o mapeamento digital de solos. Segundo IBGE (2007), os limites de densidade de amostragem recomendados para levantamentos detalhados de solos variam de 0,2 a 4 pontos.ha⁻¹. Obter indicativos concretos para amostragem em diferentes escalas é importante na busca de procedimentos padronizados para o mapeamento digital de solos, os quais já existem há mais de 50 anos para os levantamentos convencionais.

4.5.4. Sobre os algoritmos de classificação

A árvore de decisão Gini foi o algoritmo que apresentou o melhor desempenho, obtendo acurácia mais elevada para os mapas estimados em todas as densidades de amostragem. Além disso, suas regras de classificação mostram explicitamente os limites utilizados em cada subdivisão nos valores das variáveis, o que as torna compreensíveis e facilita a interpretação por parte do usuário. Esse aspecto permite avaliar a consistência das regras que descrevem as relações solo-paisagem, confrontando-as com o conhecimento prático de pedólogos com experiência local, e também serve para recuperar o conhecimento contido em mapas de solo existentes.

Nas redes neurais as relações solo-paisagem são descritas implicitamente através de pesos entre as conexões dos neurônios, inacessíveis para o usuário. Segundo Lippitt et al. (2008), esse aspecto rende a principal

crítica feita às redes neurais artificiais: o seu caráter de “caixa preta”. Todavia, o fato de não exibir regras de classificação não significa que as redes neurais devam ser descartadas, pois seu desempenho pode ser avaliado pela qualidade do resultado final obtido.

Com relação aos valores de acurácia obtidos, os resultados encontrados concordam com estudos anteriores que obtiveram desempenho superior para árvores de decisão em comparação com outros algoritmos de aprendizagem de máquina, além de gerar regras de classificação facilmente compreensíveis (Foody, 1995; Pal & Mather, 2002; Gahegan, 2003, Lippitt et al., 2008; Sarmiento et al., 2009).

Entretanto, há discordâncias com relação ao desempenho das redes neurais artificiais. Lippitt et al. (2008), por exemplo, comparando duas árvores de decisão e três redes neurais na classificação de imagens de sensoriamento remoto, obtiveram acurácia mais elevada com a árvore de decisão Gini. Entre as redes neurais o melhor desempenho foi obtido com a rede neural SOM, sendo bastante próximo ao da árvore de decisão. Esse resultado está de acordo com o esperado, já que redes neurais SOM representam uma evolução das redes neurais artificiais por se basearem em aspectos cognitivos específicos das redes neurais biológicas. No presente estudo, entretanto, o melhor desempenho entre as redes neurais foi obtido com a rede neural MLP. Essa diferença pode estar relacionada às características intrínsecas dos tipos de dados utilizados nos dois casos ou às diferenças na configuração da estrutura das redes neurais. Com relação a esta, redes neurais MLP, embora mais simples na concepção, permitem mais modificações na estrutura do que redes SOM, tanto no número de camadas ocultas quanto no número de neurônios das diferentes camadas, o que afeta o seu desempenho final.

Independente do algoritmo utilizado, é importante ter consciência que sempre haverá limitações quanto à sua capacidade de representar a realidade. Como ressaltado por Perlich et al. (2003), citando Box (1979), não se pode perder de vista que “todos os modelos estão errados, mas alguns são úteis”. O que interessa, em última análise, é encontrar um algoritmo que permita estimar os dados desejados com um adequado grau de acurácia a partir das variáveis disponíveis. Apesar do grande número de trabalhos em mapeamento digital de solos, testando diferentes métodos para estimar classes

ou propriedades de solos, ainda há poucos estudos comparando o desempenho de diferentes algoritmos de aprendizagem de máquina entre si ou com métodos paramétricos para a predição de classes de solos.

Buscando fornecer subsídios para a escolha do algoritmo de classificação mais adequado, Rogan et al. (2008) sugerem alguns critérios de avaliação. Os autores consideram importante avaliar o impacto do tamanho do conjunto de treinamento (curva de aprendizado), o efeito de variações no conjunto de treinamento, o efeito da presença de ruído no conjunto de treinamento e a interpretabilidade dos resultados do algoritmo de classificação. Critérios de escolha têm particular relevância para o mapeamento digital de solos, onde grandes variações nos conjuntos de dados de treinamento e a ocorrência de paisagens heterogêneas requerem alta capacidade de generalização por parte dos algoritmos.

5. CONCLUSÕES

Este estudo acrescenta uma contribuição para o conhecimento do desempenho de alguns algoritmos de aprendizagem de máquina na estimativa de classes de solos em escala de detalhe, tanto do ponto de vista da acurácia obtida quanto da sensibilidade à densidade de amostragem. Os resultados obtidos fornecem indicativos úteis para trabalhos futuros e podem ser resumidos nas seguintes conclusões:

- i. É possível estimar a distribuição de ordens de solos em escala de detalhe empregando algoritmos de aprendizagem de máquina e atributos do terreno derivados de MNT como variáveis preditoras;
- ii. Tomando o mesmo conjunto de variáveis preditoras, a acurácia do mapa estimado depende do algoritmo utilizado e do número de pontos de amostragem usado para o treinamento do algoritmo;
- iii. A árvore de decisão Gini apresentou o melhor desempenho, obtendo acurácia mais elevada em todos os indicadores e menor sensibilidade à densidade de amostragem;
- iv. A rede MLP apresenta potencial para predição de ordens de solos em escala de detalhe, embora necessite de um número maior de pontos amostrais para treinamento. Por permitir mais modificações na estrutura do que as outras redes neurais, ela também deve ser mais estudada para avaliar o efeito do número de neurônios e de camadas sobre o seu desempenho;
- v. Tanto a árvore de decisão quanto as redes neurais apresentam uso relativamente simples e de processamento rápido, constituindo ferramentas com potencial de aplicação prática para estimar classes de solos em áreas carentes de mapeamentos;
- vi. Embora apenas dois tenham alcançado bom desempenho, os quatro algoritmos de aprendizagem de máquina testados mostraram-se de fácil uso por estarem implementados de forma integrada em um *software* acessível e de baixo custo. Isso evitou a necessidade de converter os dados para vários formatos de vários *softwares* e mostrou que as

ferramentas disponíveis para o mapeamento digital de solos estão se tornando mais acessíveis;

- vii. Os limites de densidade de amostragem recomendados para levantamentos convencionais em escala detalhada podem ser usados como referência para dimensionar a amostragem em trabalhos de mapeamento digital de solos na mesma escala;
- viii. Os mapas digitais estimados foram produzidos em intervalo de tempo reduzido, apresentaram maior detalhamento espacial que o mapa convencional existente e são prontamente utilizáveis em SIG;
- ix. Estudos futuros devem testar outros algoritmos e também testar a inclusão de variáveis preditoras adicionais, além de comparar os resultados com dados observados em campo com vistas a avaliar o grau de aproximação dos mapas estimados com a realidade.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- AB'SABER, A. N. Conhecimentos sobre as flutuações climáticas do quaternário no Brasil. **Boletim da Sociedade Brasileira de Geografia**. São Paulo, n. 6, p. 41 - 48, 1957.
- BAILEY, N.; CLEMENTS, T.; LEE, J.T.; THOMPSON, S. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. **Journal of Environmental Management**, Amsterdam, v. 67, p. 395 - 407, 2003.
- BEHRENS, T.; FÖERSTER, H.; SCHOLTEN, T.; STEINRÜCKEN, U.; SPIES, E. GOLDSCHMITT. Digital soil mapping using artificial neural networks. **Journal of Plant Nutrition Soil Science**, v. 168, p. 21 – 33, 2005.
- BEHRENS, T.; SCHOLTEN, T. Digital soil mapping in Germany - a review. **Journal of Plant Nutrition Soil Science**, v. 169, p. 434 – 443, 2006.
- BOX, G. E. P. Robustness in the strategy of scientific model building. In LAUNER, R. L.; WILKINSON, G. N. (Ed.). **Robustness in Statistics**, New York: Academic Press, 1979. p. 201 – 236.
- BOURENNANE, H.; KING, D.; COUTURIER, A. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. **Geoderma**, Amsterdam, v. 97, p. 255 – 271, 2000.
- BUI, E. N. Soil survey as a knowledge system. **Geoderma**, Amsterdam, v. 120, p. 17-26, 2004.
- BURROUGH, P. A.; McDONNELL, R. A. **Principles of Geographical Information Systems**, Oxford: Oxford University Press, 1998.
- BURROUGH, P.A.; van GAANS, P.F.M.; MacMILLAN, R.A. High-resolution landform classification using fuzzy k-means. **Fuzzy Sets and Systems**, v. 113, p. 37 - 52, 2000.
- BRODSKY, L.; BORUVKA, L.; HAJEK, F. Object-oriented fuzzy analysis of remote sensing data for bare soil mapping. In: GLOBAL WORKSHOP ON DIGITAL SOIL MAPPING, 2. **Proceedings**. Rio de Janeiro. 2006. CD-ROM. 6 p.
- CÂMARA, G.; MEDEIROS, J. S. **Geoprocessamento para projetos ambientais**. São José dos Campos: INPE, 1996
- CAMPBELL, W. G.; MORTENSON, D. C. Ensuring the quality of geographic information system data: a practical application of quality control. **Photogrammetric Engineering and Remote Sensing**, v. 55, p. 1613 – 1618, 1989.

CARRÉ, F.; McBRATNEY, A. B.; MAYR, T.; MONTANARELLA, L. Digital soil assessments: Beyond DSM. **Geoderma**, Amsterdam, v. 142, p. 69 – 79, 2007.

CHAGAS, C. S.; FERNANDES FILHO, E. I.; VIEIRA, C. A. O.; CARVALHO JUNIOR, W. 2007. Utilização de redes neurais artificiais para predição de classes de solo em uma bacia hidrográfica no Domínio do Mar de Morros. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13. **Anais**, p. 2421 – 2428.

CHAN, J.C.; CHAN K.; YEH, A. G. Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. **Photogrammetric Engineering and Remote Sensing**, v. 67, n. 2, p. 213 - 225, 2001.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v.20, n. 1, p. 37 - 46, 1960.

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

CONTADOR, J. F. L.; MANETA, M.; SCHNABEL, S. Prediction of near-surface soil moisture at large scale by digital terrain modeling and neural networks. **Environmental Monitoring and Assessment**, v. 121, p. 213 – 232, 2006.

CRIVELANTI, R. C. **Mineração de dados para inferência da relação solo-paisagem em mapeamentos digitais de solos**. 2009, 107 p. Dissertação (Mestrado Gestão de Recursos Agroambientais) - Instituto Agrônomo de Campinas, Campinas, 2009.

DEBELLA-GILO, M.; ETZELMULLER, B.; KLAKEGG, O. Digital soil mapping using digital terrain analysis and statistical modeling integrated into GIS: Examples from Vestfold county of Norway. ScanGIS'2007. In: SCANDINAVIAN RESEARCH CONFERENCE ON GEOGRAPHICAL INFORMATION SCIENCES, 11. 2007, Norway. **Proceedings**. Norway, 2007, 274 p.

DOBOS, E.; CARRÉ, F.; HENGL, T.; REUTER, H.I.; TÓTH, G. **Digital Soil Mapping as a support to production of functional maps.**, Luxemburg: Office for Official Publications of the European Communities, 2006, 68 p. (EUR 22123 EN)

EASTMAN, J. R., 2009a. **IDRISI Taiga**: Guide to GIS and Image Processing. Clark University. 325 p., ilust.

EASTMAN, J.R. 2009b. **IDRISI Taiga** (Worcester, MA: Clark University)

EHSANI, A. H.; QUIEL, F. Application of Self-Organizing Map and SRTM data to characterize yardangs in the Lut desert, Iran. **Remote Sensing of Environment**, v.112, p. 3284 – 3294, 2008.

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

ELNAGGAR, A. A.; NOLLER, J. S. Application of Remote Sensing Data and Decision-Tree Analysis to Mapping Salt-Affected Soils over Large Areas.

Remote Sensing, v. 2, p. 151-165, 2010. Disponível em: <http://dx.doi:10.3390/rs2010151>. Acesso em: 19 de janeiro de 2010.

EMBRAPA. **Dados Meteorológicos**. Bento Gonçalves: EMBRAPA Uva e vinho. Disponível em: <http://www.cnpuv.embrapa.br/prodserv/meteorologia/bento-normais.html>. Acesso em: 23 de agosto de 2009.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 2.ed. Rio de Janeiro: Embrapa Solos, 2006, 306p.

FALCADE, I.; FLORES, C. A.; FASOLO, P. J.; PÖTTER, R. O. O Vale dos Vinhedos no contexto ambiental regional. In: FALCADE, I.; MANDELLI, F. (Org.). **Vale dos Vinhedos: caracterização geográfica da região**. Caxias do Sul: EDUCS, 1999. p. 35-40.

FILIPPI, A. M.; JENSEN, J. R. Fuzzy. Learning vector quantization for hyperspectral coastal vegetation classification. **Remote Sensing of Environment**, v. 100, p. 512 – 530, 2006.

FLORES, C. A.; FASOLO, P. J.; PÖTTER, R. O. Solos: levantamento semidetalhado. In: FALCADE, I.; MANDELLI, F. (Org.). **Vale dos Vinhedos: caracterização geográfica da região**. Caxias do Sul: EDUCS, 1999. p.87-134.

FLORES, C. A.; FASOLO, P. J.; PÖTTER, R. O. Folha SH.22-V-D-II-2. Bento Gonçalves. Região da Serra Gaúcha – RS. **Projeto Zoneamento Vitivinícola do Estado do Rio Grande do Sul**. Embrapa/UFRGS/IBRAVIN. 2006, a.

FLORES, C. A.(Org.); PÖTTER, R. O.; FASOLO, P. J.; HASENACK, H.; WEBER, E. **Levantamento semidetalhado de solos: região da Campanha – Folha Palomas, Estado do Rio Grande do Sul**. Porto Alegre: Editora da UFRGS, 2006, b.

FLORES, C. A. Comunicação pessoal. Pelotas, EMBRAPA Clima Temperado, 15 de outubro de 2006, c.

FLORES, C. A.; HASENACK, H.; WEBER, E.; SARMENTO, E. Potencial edáfico da Serra Gaúcha, Brasil para viticultura. In: CONGRESSO LATINOAMERICANO DE VITICULTURA E ENOLOGIA, 11. **Anais**. Mendoza, 2007.

FLORES, C. A. Comunicação pessoal. Pelotas, EMBRAPA Clima Temperado, 18 de março de 2009.

FLORINSKY, I.V.; EILERS, R. G.; MANNING, G.; FULLER, L.G. Prediction of soil properties by digital terrain modelling. **Environmental Modelling and Software**, v. 17, p. 295 – 311, 2002.

FOODY, G. M. Land-cover classification by an artificial neural network with ancillary information, **International Journal of Geographical Information Systems**, v. 9, n. 5, p. 527 – 542, 1995.

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

FORTES, A. B. **Compêndio de geografia geral do Rio Grande do Sul**. Porto Alegre, 1979. Sulina. 101 p.

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

GIASSON, E.; CLARKE, R. T.; INDA JUNIOR, A. V.; MERTEN, G. H.; TORNQUIST, C. G. Digital soil mapping using logistic regression on terrain parameters: A case study in Southern Brazil. **Scientia Agricola**, Piracicaba, SP, v. 63, n. 3, p. 262 - 268, 2006.

GIASSON, E.; FIGUEIREDO, S. R.; TORNQUIST, C. G.; CLARKE, R. T. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in southern Brazil. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008, p. 225-232.

GONZALEZ, J. P.; COOK, S.; OBERTHUR, T.; JARVIS, A.; BAGNELL, J. A.; DIAS, M. B. Creating Low-Cost Soil Maps for Tropical Agriculture using Gaussian Processes. **Workshop on AI in ICT for Development (ICTD)**, TWENTIETH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, Hyderabad, India, 2007.

GRAY, J.M.; HUMPHREYS, G.S.; DECKERS J.A. Relationships in soil distribution as revealed by a global soil database, **Geoderma**, Amsterdam, v. 150, p. 309 – 323, 2009.

GRINAND, C.; ARROUAYS, D.; LAROCHE, B.; MARTIN, M. P. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. **Geoderma**, Amsterdam, v. 143, p. 180 – 190, 2008.

GRIMM, R.; BEHRENS, T. Uncertainty analysis of sample locations within digital soil mapping approaches. **Geoderma**, Amsterdam, v. 155, p. 154 – 163, 2010.

HANSEN, M. K.; BROWN, D. J.; DENNISON, P. E.; GRAVES, S. A.; BRICKLEMYER, R. S. Inductively mapping expert-derived soil-landscape units within Dambo Wetland catenae using multispectral and topographic data. **Geoderma**, Amsterdam, v. 150, p. 72 – 84, 2009.

HASENACK, H.; FERRARO, L. W. Considerações sobre o clima da região de Tramandaí, RS. **Pesquisas**, n.o 22, p. 53 – 70, 1989.

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.

HENGL, T.; HEUVELINK, G.B.M. New challenges for predictive soil mapping. In: **Global Workshop on Digital Soil Mapping**. INRA: Montpellier, 2004. 9 p.

HENGL, T. **A Practical Guide to Geostatistical Mapping of Environmental Variables**. Luxembourg: JRC, 2007. 165 p.

HENGL, T.; TOOMANIAN, N.; REUTER, H. I.; MALAKOUTI, M. J.. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, Amsterdam v. 140, p. 417 – 427, 2007.

HEMPEL, J. W.; HAMMER, R. D.; MOORE, A. C.; BELL, J. C.; THOMPSON, J. A.; GOLDEN, M. L. Challenges to Digital Soil Mapping. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008, p. 81 – 90.

HENDERSON, B.L.; BUI, E.N.; MORAN, C.J.; SIMON, D.A.P. Australia-wide predictions of soil properties using decision trees. **Geoderma**, Amsterdam, v.124, p. 383 – 398, 2005.

HEUVELINK, G.B.M.; WEBSTER, R. Modelling soil variation: past, present, and future. **Geoderma**, Amsterdam, v. 100, p. 269–301, 2001.

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

HOFFMANN, A.; GUERRA, C. C.; TONIETTO, J., MELLO, L. M. R.; HOFF, R.; CAMARGO, U. A.; FALCADE, I.; HASENACK, H.; SARMENTO, E. C.; WEBER, E. Serra Gaúcha - demarcação da zona de produção vitivinícola. Bento Gonçalves: Embrapa Uva e Vinho, 2005. **Nota Técnica**. 25 p.

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

IBGE. **Folha SH.22 Porto Alegre e parte das Folhas SH.21 Uruguaiana e SI.22 Lagoa Mirim**. Rio de Janeiro, IBGE, 1986. CD-ROM. (Levantamento de Recursos Naturais, v. 33)

IBGE. **Geografia do Brasil - Região Sul**. Rio de Janeiro: IBGE, 1990.

IBGE. **Manual técnico de pedologia**.2.ed. Rio de Janeiro:IBGE, 2007, 300 p.

IPPOLITI, G.; COSTA, L.M. da; GAGGERO, M.R.; SCHAEFFER, C. E.; FERNANDEZ FILHO, E.I. 2003. Uso de geoprocessamento para o levantamento preliminar de solos de uma microbacia da zona da mata (MG). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 11. **Anais**. Belo Horizonte. 2003. p. 153-160.

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

JENSON, S.; DOMINGUE, J. Extracting topographic structure from digital elevation data for geographic information system analysis. **Photogrammetric Engineering and Remote Sensing**, v. 54, n.11, p. 1593 - 1600, 1988.

JOHNSTON, D. M.; TIMLIN, D. Spatial data accuracy and quality assessment for environmental management. In: HEUVELINK, G. B. M.; LEMMENS, M. J. P.

M. (Ed.). International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 4. **Proceedings**, 2000. p. 325–328. Delft: Delft University Press.

KANELLOPOULOS, I.; WILKINSON, G. G. Strategies and best practice for neural network image classification, **International Journal of Remote Sensing**, v. 18, n. 4, p. 711 – 725, 1997.

KAVZOGLU, T. Increasing the accuracy of neural network classification using refined training data. **Environmental Modelling & Software**, v. 24, p. 850 – 858, 2009.

KOHONEN, T. The Self-Organizing Map, IEEE, 78. **Proceedings**, p. 1464 – 1480. 1990.

LAGACHERIE, P., MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: LAGACHERIE, P.; McBRATNEY, A. B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: an introductory perspective**. Elsevier: Amsterdam, 2007, p. 3 - 24. (Developments in soil science, vol. 31)

LAGACHERIE, P. Digital Soil Mapping: A State of the Art. In.: HARTEMINK, A. E.; McBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.), **Digital Soil Mapping with Limited Data**. New York: Springer, 2008, p. 3 - 14.

LANDIS, J., KOCH, G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, p. 159 – 174, 1977.

LEVINE, E. R.; KIMES, D. S.; SIGILLITO, V. G. Classifying soil structure using neural networks. **Ecological Modelling**, v. 92, p. 101 - 108, 1996.

LI, Z.; ZHU, Q.; GOLD, C. **Digital Terrain Modeling: Principles and Methodology**. Boca Raton: CRC Press, 2005.

LI, Z.; EASTMAN, J. R. The nature and classification of unlabelled neurons in the use of Kohonen's self-organizing map for supervised classification, **Transactions in GIS**, v. 10, n.4, p. 599 – 613, 2006.

LIPPITT, C. D.; ROGAN, J.; LI, Z.; EASTMAN, J. R.; JONES, T. G. Mapping Selective Logging in Mixed Deciduous Forest: A Comparison of Machine Learning Algorithms. **Photogrammetric Engineering and Remote Sensing**, v. 74, n. 10, p. 1201 – 1211, 2008.

LIU, J.; ZHU, A. Mapping with Words: A New Approach to Automated Digital Soil Survey. **International Journal of Intelligent Systems**, v. 24, p. 293 – 311, 2009.

LOPEZ, L. R. **Pedologia quantitativa: espectrometria VIS-NIR-SWIR e mapeamento digital de solos**. 2009, 171 p. Dissertação (Mestrado). Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2009.

LOPEZ, L. R.; DEMATTÊ, J. A. M. Desenvolvendo uma metodologia pedométrica integrada para predição digital de mapas detalhados de solo. In:

SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, **Anais**. INPE, Natal. 2009. p. 355 – 362.

LUOTO, M.; HJORT, J. Evaluation of current statistical approaches for predictive geomorphological mapping. **Geomorphology**, v. 67, p. 299 – 315, 2005.

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

MANNAN, B.; ROY, J.; RAY, A. K. Fuzzy ARTMAP supervised classification of remotely-sensed images, **International Journal of Remote Sensing**, v. 19, n. 4, p. 767 – 774, 1998.

MAS, J. F.; FLORES, J. J. The application of artificial neural networks to the analysis of remotely sensed data. **International Journal of Remote Sensing**, v. 29, n. 3, p. 617 – 663, 2008.

McBRATNEY, A. B.; ODEH, I. O. A.; BISHOP, T. F.A.; DUNBAR, M. S. An overview of pedometric techniques for use in soil survey. **Geoderma**, Amsterdam, v. 97, p. 293 – 327, 2000.

McBRATNEY, A.B.; MENDONÇA-SANTOS, M.L.; MINASNY, B. On Digital Soil Mapping. **Geoderma**, Amsterdam, v. 117, p. 3 - 52, 2003.

McKENZIE, N.J.; RYAN, P.J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, Amsterdam, v. 89, p. 67 - 94, 1999.

MENDONÇA-SANTOS, M.L.; GUENAT, C.; BOUZELBOUDJEN, M.; GOLAY, F. Three-dimensional GIS cartography applied to the study of the spatial variation of soil horizons in a Swiss floodplain. **Geoderma**, Amsterdam, v. 97, p. 351 – 366, 2000.

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

MINASNY, B., McBRATNEY, A. B. The neuro-m method for fitting neural network parametric pedotransfer functions. **Soil Science Society of America Journal**, Madison, v.66, p. 352 – 361, 2002.

MORENO, J. A. **Clima do Rio Grande do Sul**. Porto Alegre: Secretaria da Agricultura, 1961. 42p.

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

MOORE, I. D.; GESSLER, P. E.; NIELSEN, G. A.; PETERSON G. A. Soil attribute prediction using terrain analysis. **Soil Science Society American Journal**, Madison, v. 57, p. 443 - 452, 1993.

MORRIS, D. K.; STIENHARDT, G. C.; NIELSEN, R. L.; HOSTETTER, W.; HALEY, S.; STRUBEN, G. R. Using GPS, GIS, and Remote Sensing as a Soil Mapping Tool. INTERNATIONAL CONFERENCE ON PRECISION AGRICULTURE, 5. **Proceedings**. Bloomington, MN. Center for Precision Agriculture, University of Minnesota, St. Paul, MN. 2000.

NIMER E. **Climatologia do Brasil**. Rio de Janeiro: IBGE, 1989. 421p.

PAL, M.; MATHER, P. M. A Comparison of Decision Tree and Backpropagation Neural Network Classifiers for Land Use Classification. INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2. **Proceedings**, vol. 1. IEEE International, Nottingham. p. 503 – 505. 2002.

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

PENNOCK, D. J. Terrain attributes, landform segmentation, and soil redistribution. **Soil & Tillage Research**, v. 69, p. 15 – 26, 2003.

PERLICH, C.; PROVOST, F.; SIMONOFF, J. S.. Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. **Journal of Machine Learning Research**, v. 4, p. 211 - 255, 2003.

PONTIUS, R. G. Quantification Error versus Location Error in Comparison of Categorical Maps. **Photogrammetric Engineering and Remote Sensing**, v. 66, n.8, p. 1011 - 1016, 2000.

PROTAS, J. F. S; CAMARGO, U. A.; MELLO, L. M. R. **A vitivinicultura brasileira: realidade e perspectivas**. Embrapa Uva e Vinho, 2004. Disponível em: <<http://www.embrapa.cnpuv.embrapa.br/publica/artigos/vitivinicultura/>>. Acesso em: 5 de março de 2007.

QI, F.; ZHU, A. X. Knowledge discovery from soil maps using inductive learning. **International Journal of Geographical Information Science**, v. 17, n.8, p. 771 – 795, 2003.

QI, F. Knowledge discovery from area-class resource maps: data preprocessing for noise reduction. **Transactions in GIS**, v. 8, n. 3, p. 297 – 308, 2004.

QI, F.; ZHU, A. X.; HARROWER, M.; BURT, J.E.. Fuzzy soil mapping based on prototype category theory. **Geoderma**, Amsterdam, v. 136, p. 774 – 787, 2006.

RICHARDS, J. A.; JIA, X. **Remote Sensing Digital Image Analysis**. New York: Springer. 1999.

ROGAN, J.; MILLER, J.; STOW, D. A.; FRANKLIN, J.; LEVIEN, L.; FISCHER, C. Land-cover change mapping in California using classification trees with Landsat TM and ancillary data, **Photogrammetric Engineering & Remote Sensing**, v.69, n.7, p. 793 – 804, 2003.

ROGAN, J.; FRANKLIN, J.; STOW, D.; MILLER, J.; ROBERTS, D.A.; WOODCOCK, C. Mapping land cover modifications over large areas: A

comparison of machine learning algorithms. **Remote Sensing of Environment**, v. 112, n. 5, p. 2272 – 2283, 2008.

ROMANO, N.; CHIRICO, G. B. The role of terrain analysis in using and developing pedotransfer functions. PACHEPSKY, Y; RAWLS, W.J. (Ed.) **Development of pedotransfer functions in soil hydrology**. Amsterdam: Elsevier, 2004. p. 273 – 294. (Developments in Soil Science, v.30)

ROSSITER, D. G. Assessing the thematic accuracy of area-class soil maps. ITC: Enschede, 2001. Disponível em: <http://www.itc.nl/~rossiter/docs/AccuracySoilMaps_Preprint31Aug2001.pdf> Acesso em: 20 de abril de 2009.

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

ROSSITER, D. G. Digital soil resource inventories: status and prospects. **Soil Use and Management**, v. 20, p. 296 – 301, 2004, b.

ROSSITER, D. G. **Digital soil mapping**: Towards a multiple-use Soil Information System. Semana de la Geomática. Bogotá, 2005. Disponível em: <<http://www.itc.nl/personal/rossiter>>. Acesso em: 24 de setembro de 2009.

ROSSITER, D. G. Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures. In: HARTEMINK, A. E.; McBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008, p. 69 - 80.

SANCHEZ, P. A.; AHAMED, S.; CARRÉ, F.; HARTEMINK, A. E.; HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; McBRATNEY, A. B.; MCKENZIE, N. J.; MENDONÇA-SANTOS, M. L.; MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C. A.; SACHS, J. D.; SHEPHERD, K. D.; VÅGEN, T. G.; VANLAUWE, B.; WALSH, M. G.; WINOWIECKI, L. A.; ZHANG, G. L. Digital Soil Map of the World. **Science**, v. 35, p. 681 – 682, 2009.

SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C.; ANJOS, L.H.C. **Manual de descrição e coleta de solo no campo**. 5. ed. Viçosa: Sociedade Brasileira de Ciência do Solo, 2005. 92 p.

SARMENTO, E. C.; WEBER, E. J.; HASENACK, H.; TONIETTO, J.; MANDELLI, F. Topographic modeling with GIS at Serra Gaúcha, Brazil: elements to study viticultural terroir. In: TERROIR VITICOLES 2006. CONGRES INTERNACIONAL, 6. Bordeaux: **Vigne et vin Publications Internationales**, v.1, p. 365 - 372. 2006.

SARMENTO, E. C.; FLORES, C. A.; WEBER, E.; HASENACK, H. & PÖTTER, R. O. Sistema de informação geográfica como apoio ao levantamento detalhado de solos do Vale dos Vinhedos. **Revista Brasileira de Ciência do Solo**. v. 32, p. 2795 - 2803, 2008.

SARMENTO, E. C.; WEBER, E.; GIASSON, E.; FLORES, C. A.; HASENACK, H. Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos do Vale dos Vinhedos, RS, Brasil. In:

CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 32. **Livro de resumos**. Fortaleza, 2009. CD-ROM.

SCULL, P.; FRANKLIN, J.; CHADWICK, O. A. McARTHUR, D. 2003. Predictive soil mapping: a review. **Progress in Physical Geography**, v. 27, n. 2, p. 171 – 197, 2003.

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

SHI X.; ZHU, A. X.; BURT, J. E.; QI, F.; SIMONSON, D. A Case-based Reasoning Approach to Fuzzy Soil Mapping. **Soil Science Society of America Journal**, Madison. v. 68, p. 885 – 894, 2004.

STRECK, E. V.; KÄMPF, N.; DALMOLIN, R. S. D.; KLAMT, E.; NASCIMENTO, P. C.; SCHNEIDER, P.; GIASSON, E.; PINTO, L. F. S. **Solos do Rio Grande do Sul**. 2. ed. Porto Alegre: EMATER/RS, 2008. 222 p.

TONIETTO J.; MANDELLI, F. Como organizar, promover y reconocer regiones de excelencia de producción de vinos: Una experiencia de Brasil en Indicadores Geográficos. In: SEMINARIO INTERNACIONAL DE VITIVINICULTURA, 2. **Anais**. 2005. Ensenada, Mexico. Inifap, Mexico. p. 7 - 19.

WALTER, C.; LAGACHERIE, P.; FOLLAIN, S. Integrating pedological knowledge into digital soil mapping In: LAGACHERIE, P.; MCBRATNEY, A.; VOLTZ, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 281 - 300.

WEBER, E.; HASENACK, H.; FLORES, C.A.; PÖTTER, R.O.; FASOLO, P.J. GIS as a support to soil mapping in southern Brazil. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008, p. 103 - 112.

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

WONG, G. Y.; MASON W. M. The hierarchical logistic regression model for multilevel analysis. **Journal of the American Statistical Association**, Alexandria, VA, v. 80, p. 513 - 524, 1985.

XU, M.; WATANACHATURAPORN, P.; VARSHNEY, P. K.; ARORA, M. K. Decision tree regression for soft classification of remote sensing data. **Remote Sensing of Environment**, v. 97, p. 322 – 336, 2005.

ZAMBON, M.; LAWRENCE, R.; BUNN, A.; POWELL, S. Effect of alternative splitting rules on image processing using classification tree analysis. **Photogrammetric Engineering and Remote Sensing**, v. 72, n. 1, p. 25 – 30, 2006.

ZHOU, B.; ZHANG, X.; WANG, R. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. **Journal of Zhejiang University Science**, v. 5, n. 7, p. 782 - 795, 2004.

ZHU, A. X.; BAND, L. E.; DUTTON, B.; NIMLOS, T. J. Automated soil inference under fuzzy logic. **Ecological Modelling**, v. 90, p. 123 - 145, 1996.

ZHU, A. X. A similarity model for representing soil spatial information. **Geoderma**, Amsterdam, v. 77, p. 217 – 242, 1997.

ZHU, A. X. Mapping soil landscape as spatial continua: the neural network approach. **Water Resources Research**, v. 36, p. 663 - 677, 2000.

ZHU, A. X.; HUDSON, B.; BURT, J.; LUBICH, K.; SIMONSON, D. Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic. **Soil Science Society of America Journal**., v. 65, p. 1463 – 1472, 2001.

ZHU, A. X.; YANG, L.; LI, B.; QIN, C.; ENGLISH, E.; BURT, J. E.; ZHOU, C. Purposive Sampling for Digital Soil Mapping for Areas with Limited Data. In: HARTEMINK, A. E.; MCBRATNEY, A.; MENDONÇA-SANTOS, M. L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008. p. 233 – 245.

ZIADAT, F. M. Analyzing Digital Terrain Attributes to Predict Soil Attributes for a Relatively Large Area. **Soil Science Society of America Journal**. v. 69, p. 1590 – 1599, 2005.