

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

**PREDIÇÃO DE CLASSES DE SOLOS EM DIFERENTES ESCALAS NA
SERRA GAÚCHA USANDO MAPEAMENTO DIGITAL DE SOLOS A PARTIR
DE DADOS LEGADOS**

Eliana Casco Sarmento

(Tese)

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

**PREDIÇÃO DE CLASSES DE SOLOS EM DIFERENTES ESCALAS NA
SERRA GAÚCHA USANDO MAPEAMENTO DIGITAL DE SOLOS A PARTIR
DE DADOS LEGADOS**

Eliana Casco Sarmiento

Engenheira Agrônoma e Bióloga (UFRGS)
Mestre em Ciência do Solo (UFRGS)

Tese apresentada como
um dos requisitos à obtenção do
grau de Doutor em Ciência do Solo

Porto Alegre (RS) Brasil
Abril de 2015.

CIP - Catalogação na Publicação

Sarmiento, Eliana Casco

Predição de classes de solos em diferentes escalas na Serra Gaúcha usando mapeamento digital de solos a partir de dados legados / Eliana Casco Sarmiento. -- 2015.

116 f.

Orientador: Elvio Giasson.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Programa de Pós-Graduação em Ciência do Solo, Porto Alegre, BR-RS, 2015.

1. Pedologia. 2. Mapeamento digital. 3. Predição de classes. 4. Extrapolação. 5. Desagregação. I. Giasson, Elvio, orient. II. Título.

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

ELIANA CASCO SARMENTO

PREDIÇÃO DE CLASSES DE SOLOS EM DIFERENTES ESCALAS NA
SERRA GAÚCHA USANDO MAPEAMENTO DIGITAL DE SOLOS A PARTIR
DE DADOS LEGADOS

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência do Solo da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul, como requisito para a obtenção do título de Doutora em Ciência do Solo.

Prof. Orientador: Elvio Giasson

Aprovada em 29 de abril de 2015
Homologada em 08 de julho de 2015

Banca examinadora:

Prof. Carlos Gustavo Tornquist
UFRGS

Dr. Ivan Luiz Zilli Bacic
EPAGRI-SC
(através de videoconferência)

Dr. Egon Klamt

AGRADECIMENTOS

Aos meus pais, Nelson (*in memoriam*) e Paulina, por tudo quanto representam para mim;

Ao meu orientador, professor Élvio Giasson, por acreditar na proposta de pesquisa quando ela era pouco mais que uma idéia, pela orientação e contribuição em temas essenciais, bem como pela confiança e liberdade proporcionada na condução das atividades;

À UFRGS, onde cursei minhas duas graduações, e em especial ao PPG em Ciência do Solo, onde também cursei meu mestrado, seus professores, colegas e funcionários, por tudo que pude aprender e vivenciar nesses anos;

À CAPES, pela bolsa de estudos que possibilitou total dedicação ao curso de doutorado e a elaboração desta tese, bem como pela disponibilização do portal de periódicos;

Ao CNPq, à FINEP e à EMBRAPA, pelo financiamento de projetos que, embora não diretamente ligados à pesquisa desta tese, propiciaram dados nela utilizados;

Aos professores Carlos Gustavo Tornquist e Paulo César do Nascimento, do PPG em Ciência do Solo, e ao pesquisador Ivan Luiz Zilli Bacic, da EPAGRI, pelas contribuições ao aprimoramento do projeto de pesquisa na banca de qualificação;

Ao pesquisador Carlos Alberto Flores, da Embrapa Clima Temperado, que desde a participação como estudante no famoso Brasil 1973 mantém a mesma empolgação pela pedologia, pela generosidade em compartilhar sua experiência e seu inigualável bom humor, pelo auxílio em inesquecíveis expedições em campo, e pela gentileza em ceder seus dados sempre com a mente aberta ao novo;

Ao professor Heinrich Hasenack, do Centro de Ecologia da UFRGS, pela amizade, estímulo, sugestões e oportunidade de ter participado em diversos projetos, alguns dos quais supriram dados fundamentais à pesquisa desta tese;

Ao pesquisador Jorge Tonietto, da Embrapa Uva e Vinho, pela sensibilidade à importância do solo ao liderar vários projetos de indicações geográficas, propiciando recursos financeiros inestimáveis para novos levantamentos e testes de metodologias;

Ao professor David Goodrich Rossiter, da *Cornell University*, EUA, colaborador do *ISRIC - World Soil Information*, Holanda, pelo interesse por um estudo sobre escala efetiva apresentado no Congresso Brasileiro de Ciência do Solo de 2013, e pelo subsequente encorajamento e sugestões de ampliação e aprofundamento que resultaram no artigo sobre avaliação de mapas legados do Capítulo II;

Ao professor James (Jim) A. Thompson, da *West Virginia University*, EUA, pela prestatividade e valiosas sugestões sobre os primeiros passos para a desagregação de mapas convencionais de solos na situação de poucos dados descritivos;

Ao consultor Robert (Bob) A. MacMillan, da *LandMapper Environmental Solutions Inc.*, Canadá, pela gentil cedência do programa LandMapr© e pela disposição permanente em auxiliar e esclarecer dúvidas no cálculo de classes de geoformas, as quais foram cruciais na metodologia de desagregação apresentada no Capítulo III;

Ao professor Alfred Hartemink, da *University of Wisconsin-Madison*, EUA, pelas relevantes sugestões quanto à seleção e organização do emaranhado de resultados para a elaboração do artigo sobre extrapolação referente ao Capítulo IV;

Aos meus amigos e a muitas outras pessoas que não foram nominalmente citadas, mas que de alguma forma contribuíram para a obtenção desta tese;

E em especial a meu companheiro Eliseu, pela compreensão, amor e paciência que foram imprescindíveis durante todo o doutorado, por estar sempre ao meu lado, me ajudar e me incentivar, muito obrigado.

PREDIÇÃO DE CLASSES DE SOLOS EM DIFERENTES ESCALAS NA SERRA GAÚCHA USANDO MAPEAMENTO DIGITAL DE SOLOS A PARTIR DE DADOS LEGADOS^{1/}

Autor: Eliana Casco Sarmiento
Orientador: Prof. Élvio Giasson

RESUMO

Dados de levantamentos convencionais de solos têm sido muito usados como entrada em mapeamento digital de solos (MDS), frequentemente sem considerar a qualidade original. O objetivo desta pesquisa foi avaliar metodologias e ferramentas de MDS para prever classes de solo a partir de levantamentos existentes. Um conjunto diversificado de mapas convencionais foi caracterizado com indicadores computados a partir das unidades de mapeamento (UM) e da legenda. Um procedimento de desagregação foi testado em um mapa semidetalhado, definindo-se áreas típicas das classes por UM a partir do relatório de levantamento e expedições em campo. 10% delas foram empregadas para treinar uma árvore de decisão (AD) e prever classes em toda a área, usando 21 covariáveis. A predição em locais não mapeados (extrapolação) foi avaliada em um mapa detalhado, no qual três subáreas representativas foram usadas para gerar pontos com duas densidades e dois esquemas de distribuição espacial. Os pontos foram usados para treinar uma AD e prever ordens e subordens em áreas não amostradas, com duas resoluções espaciais e variação do número e combinação de covariáveis. A desagregação possibilitou incrementar consideravelmente o detalhamento do mapa convencional de solos, sem perda de acurácia. A extrapolação possibilitou mapear áreas não visitadas com alta resolução espacial e acurácia equivalente à área de referência, notadamente com amostragem aleatória estratificada. Em ambos, a acurácia foi limitada pela qualidade dos dados legados, em geral inferior à presumida, o que demonstra a necessidade de levá-la em conta ao desenvolver e calibrar modelos de predição em MDS.

1/ Tese de Doutorado em Ciência do Solo. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (116 p.) Abril, 2015. Trabalho realizado com apoio financeiro da CAPES.

PREDICTION OF SOIL CLASSES AT DIFFERENT SCALES IN SERRA GAÚCHA USING DIGITAL SOIL MAPPING FROM LEGACY DATA^{2/}

Author: Eliana Casco Sarmento
Adviser: Prof. Élvio Giasson

ABSTRACT

Conventional soil survey data have been widely used as inputs for digital soil mapping (DSM), often without regard to their original quality. The objective of this research was to evaluate DSM methodologies and tools for predicting soil classes from existing soil surveys. A diversified set of conventional soil maps was characterized with indicators computed from mapping units (MU) and legend. A disaggregation procedure was tested on a semidetained soil map, by defining typical areas of soil classes in each MU based on the soil survey report and field work. 10% of these areas were employed to train a decision tree (DT) and to predict soil classes throughout the area using 21 soil covariates. The prediction in unmapped areas (extrapolation) was evaluated on a detailed soil map, for which three representative subareas were used to generate points using two sampling densities and two spatial distribution schemes. These sampling points were used to train a DT and then to predict soil orders and suborders at non sampled areas, using two spatial resolutions and varying number and combination of covariates. Disaggregation allowed to increase considerably the details of the conventional soil map, without loss of accuracy. Extrapolation allowed mapping non visited areas with high spatial resolution and accuracy equivalent to the reference area, especially with stratified random sampling. In both cases, accuracy was limited by the quality of legacy soil data, in general lower than presumed, what demonstrates the need to take it into account when building and calibrating prediction models in DSM.

^{2/} Doctoral Thesis in Soil Science. Programa de Pós-Graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul. Porto Alegre. (116 p.) April, 2015. Work carried out with financial support from CAPES.

SUMÁRIO

	Página
1. INTRODUÇÃO GERAL.....	1
1.1. Objetivo e hipótese.....	3
1.2. Estrutura da tese	3
2. CAPÍTULO I – O SOLO: DA OBSERVAÇÃO EMPÍRICA AO MAPEAMENTO DIGITAL.....	5
2.1. O estudo do solo como ciência.....	5
2.1.1. Gênese dos solos	7
2.1.2. Morfologia e classificação.....	9
2.1.3. Mapeamento.....	12
2.2. Levantamentos de solos.....	14
2.3. Mapeamento digital de solos.....	17
2.3.1. Fatores <i>s.c.o.r.p.a.n</i> e atributos do terreno.....	19
2.3.2. A importância de dados legados de solos	22
2.3.3. Avaliação da qualidade de mapas digitais de solos.....	25
2.3.3.1. Validação com dados de campo	25
2.3.3.2. Outras formas de validação	27
3. CAPÍTULO II - CARACTERIZAÇÃO DE MAPAS LEGADOS DE SOLOS: USO DE INDICADORES EM MAPAS COM DIFERENTES ESCALAS NO RIO GRANDE DO SUL	29
3.1. Introdução.....	29
3.2. Material e métodos	32
3.3. Resultados e discussão.....	35
3.4. Conclusões.....	44
4. CAPÍTULO III – APERFEIÇOAMENTO DE MAPAS LEGADOS DE SOLOS POR DESAGREGAÇÃO DE POLÍGONOS USANDO CONHECIMENTO ESPECIALISTA E TÉCNICAS DE MAPEAMENTO DIGITAL.....	45
4.1. Introdução.....	45
4.2. Material e métodos	48
4.2.1. Área de estudo	48

4.2.2. Material utilizado.....	49
4.2.3. Metodologia	50
4.3. Resultados e discussão.....	59
4.3.1. Fração, número e área das classes desagregadas	59
4.3.2. Validação com <i>pedons</i> independentes	63
4.3.3. Análise visual.....	68
4.3.4. Considerações finais	72
4.4. Conclusões.....	74
5. CAPÍTULO IV – PREDIÇÃO DE CLASSES DE SOLOS COM ALTA RESOLUÇÃO ESPACIAL EM ÁREAS NÃO AMOSTRADAS A PARTIR DE UM MAPA DETALHADO EXISTENTE	75
5.1. Introdução.....	75
5.2. Material e métodos	77
5.2.1. Área de estudo	77
5.2.2. Material utilizado.....	79
5.2.3. Metodologia	79
5.3. Resultados e discussão.....	86
5.3.1. Considerações finais	99
5.4. Conclusões.....	101
6. CONCLUSÕES GERAIS.....	102
7. CONSIDERAÇÕES FINAIS	104
8. REFERÊNCIAS.....	105

RELAÇÃO DE TABELAS

Página

Tabela 1. Identificação, local, área mapeada, escala nominal e fonte de mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.....	33
Tabela 2. Escala nominal, número de polígonos, área média dos polígonos (ASD), área mínima mapeável (MLD), índice de máxima redução (IMR), escala efetiva, número de polígonos menores que a área mínima mapeável (Políg.<MLD) e parâmetros estatísticos do índice de complexidade de forma (S) para mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.....	38
Tabela 3. Escala nominal, classes taxonômicas em diferentes níveis de classificação (N1 = Ordem, N2 = Subordem, N3 = Grande Grupo, N4 = Subgrupo), tipos de solos, unidades de mapeamento e fração da área mapeada com dados não uniformes (PND) em mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.....	39
Tabela 4. Exemplos de regras para individualização da ocorrência típica das classes de solo em três UM combinadas e em uma UM simples, a partir de classes de declividade e de geoformas (LCR: level crest; DSH: divergent shoulder; BSL: back slope; DBS: divergent back slope; CBS: convergent back slope).....	56
Tabela 5. Covariáveis explanatórias e fatores de formação dos solos representados pelas mesmas na Serra Gaúcha (C: clima, R: relevo, P: material de origem, O: organismos, T: tempo).....	57
Tabela 6. Medidas de acurácia dos mapas desagregados em relação a 233 <i>pedons</i> independentes com verdade de campo.....	65
Tabela 7. Covariáveis explanatórias e fatores de formação dos solos representados pelas mesmas no Vale dos Vinhedos (C: clima, R: relevo, P: material de origem, T: tempo).....	83
Tabela 8. Comparação da EG média de mapas de classes de solo em nível de ordem e subordem, preditos a partir de covariáveis com duas resoluções espaciais e pontos amostrais gerados segundo dois esquemas de distribuição espacial e duas densidades de amostragem.....	89
Tabela 9. Indicadores de acurácia do melhor mapa predito e média de todos os mapas preditos com classes taxonômicas em nível de ordem na área total mapeada, dentro da área de treinamento e fora da área de treinamento.....	90
Tabela 10. Indicadores de acurácia do melhor mapa predito e média de todos os mapas preditos com classes taxonômicas em nível de subordem na área total mapeada, dentro da área de treinamento e fora da área de treinamento.....	90

Tabela 11. Indicadores de acurácia do mapa convencional de solos do Vale dos Vinhedos para classes taxonômicas em nível de ordem e subordem, determinados a partir de 130 <i>pedons</i> independentes coletados em campo.....	92
Tabela 12. Área ocupada e estatísticas de confiança computadas para o melhor mapa de classes preditas em nível de ordem.	98
Tabela 13. Área ocupada e estatísticas de confiança computadas para o melhor mapa de classes preditas em nível de subordem.	99

RELAÇÃO DE FIGURAS

Página

Figura 1. Estratégias para desenvolvimento de modelos preditivos a partir de dados legados de solos (Minasny & McBratney, 2010, adaptado).....	23
Figura 2. Localização de mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000 (1 = Vale dos Vinhedos; 2 = Município de Porto Alegre; 3 = Folha Palomas; 4 = Folha Pinheiro Machado; 5 = Folha Encruzilhada do Sul; 6 = bacia do rio Gravataí; 7 = Serra Gaúcha; 8 e 9 = Rio Grande do Sul).....	34
Figura 3. Localização da folha Bento Gonçalves no Rio Grande do Sul.....	48
Figura 4. Fluxograma da metodologia utilizada na desagregação de unidades de mapeamento de solos na Serra Gaúcha.	51
Figura 5. (a) Fotografia da paisagem adquirida em um local visitado em campo; (b) ponto de vista aproximadamente igual no Google Earth, com sobreposição dos limites das UM (linhas pretas) às imagens de satélite; (c) às classes de declividade; e (d) às classes de geformas.	54
Figura 6. Exemplos da tabulação cruzada de classes de declividade e de geformas, com a identificação da ocorrência típica das classes de solos em: (a) uma UM combinada com três classes; (b) uma UM simples. As letras e números abaixo dos quadros indicam as componentes e as respectivas frações na UM (LCR: level crest; DSH: divergent shoulder; UDE: upper depression; BSL: back slope; DBS: divergent back slope; CBS: convergent back slope; TER: terrace; SAD: saddle; MDE: mid slope depression; FSL: foot slope; TSL: toe slope; FAN: lower slope fan; LSM: lower slope mound; LLS: level lower slope; DEP: depression).	55
Figura 7. Relação entre a fração das componentes de cada UM após a desagregação e a fração declarada no relatório de levantamento: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.....	60
Figura 8. Diferença entre a fração das componentes de cada UM após a desagregação e sua fração declarada no relatório (erro absoluto) em relação à área das classes de solo na UM: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.	61
Figura 9. Relação entre a área total das classes de solo desagregadas e a área total calculada com base no relatório de levantamento: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.	62
Figura 10. Localização dos 233 <i>pedons</i> independentes usados para a validação dos mapas desagregados.....	64

Figura 11. Corte de estrada na área de estudo mostrando a variação espacial dos solos ao longo de uma distância que é inferior à resolução espacial utilizada (30m).	68
Figura 12. Detalhe de uma porção da área de estudo: (a) UM combinadas no mapa convencional original; (b) classes desagregadas restringindo a predição pelo limite de cada UM; (c) classes desagregadas sem restringir a predição pelo limite de cada UM (Os limites das UM foram mantidos nos dois últimos para ilustrar a distribuição espacial das classes em relação aos polígonos do mapa original).	70
Figura 13. Visão 3D no Google Earth com o resultado da avaliação da aptidão dos solos à viticultura (Flores et al., 2007d): (a) nas UM do mapa convencional, com base na classe de solo dominante; (b) nas classes desagregadas restringindo pelos limites das UM; (c) nas classes desagregadas sem restringir pelos limites das UM; (d) imagem ilustrando o uso do solo, mais coerente com as classes desagregadas do que os limites das UM (linhas pretas).	71
Figura 14. (a) Localização do Vale dos Vinhedos no estado do Rio Grande do Sul e; (b) limites do Vale dos Vinhedos.	78
Figura 15. Fluxograma da metodologia utilizada na predição de classes de solos em áreas não amostradas no Vale dos Vinhedos.	80
Figura 16. Subáreas representativas do Vale dos Vinhedos utilizadas para obtenção de amostras de treinamento (1 = área de maior altitude e relevo mais suave; 2 = área de altitude intermediária e relevo heterogêneo; 3 = área de menor altitude e relevo acidentado).	81
Figura 17. Localização de 237 <i>pedons</i> existentes (perfis e pontos de observação), usados para a validação dos mapas preditos, e de 130 <i>pedons</i> independentes coletados em campo para a validação do mapa convencional.	85
Figura 18. Exatidão geral da predição de classes de solo em nível de ordem na área total, dentro da área de treinamento e fora da área de treinamento, na resolução espacial de 5 m com amostragem (a) aleatória e (b) aleatória estratificada; e na resolução espacial de 10 m com amostragem (c) aleatória e (d) aleatória estratificada.	87
Figura 19. Exatidão geral da predição de classes de solo em nível de subordem na área total, dentro da área de treinamento e fora da área de treinamento, na resolução espacial de 5 m com amostragem (a) aleatória e (b) aleatória estratificada; e na resolução espacial de 10 m com amostragem (c) aleatória e (d) aleatória estratificada.	88
Figura 20. Exemplos de discordâncias observadas entre as classes de solo no mapa convencional e em pontos visitados em campo com distribuição espacial aleatória (o texto em vermelho corresponde à classe identificada em campo em cada ponto; o	

<p>texto em preto corresponde à classe das UM nas imediações dos pontos).</p> <p>Figura 21. Relação entre a área total das classes de solo no mapa convencional e no melhor mapa predito, considerando a classificação taxonômica em nível de (a) ordem; e (b) de subordem.</p> <p>Figura 22. (a) Melhor mapa de classes preditas em nível de ordem e; (b) confiança da predição em cada <i>pixel</i>.</p> <p>Figura 23. (a) Melhor mapa de classes preditas em nível de subordem e; (b) confiança da predição em cada <i>pixel</i>.</p>	<p>93</p> <p>94</p> <p>96</p> <p>97</p>
---	---

RELAÇÃO DE ABREVIATURAS

AM	Área Total Mapeada
AND	Área com dados não uniformes
APL	Arranjos Produtivos Locais
ASD	Tamanho médio dos polígonos delimitados
CLI	<i>Canada Land Inventory</i>
DQ	Discordância de quantidade
DA	Discordância de alocação
DSG	Diretoria de Serviço Geográfico
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ESN	Divisor da escala efetiva
EG	Exatidão Global
FAO	<i>Food and Agriculture Organization</i>
FINEP	Financiadora de Estudos e Projetos
GPS	<i>Global Positioning System</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IG	Índicações Geográficas
IMR	Índice de Máxima Redução
INDE	Infraestrutura Nacional de Dados Espaciais
KML	<i>Keyhole Markup Language</i>
MDE	Modelo Digital de Elevação
MLD	Área mínima mapeável
MDS	Mapeamento Digital de Solos
MDT	Modelo Digital do Terreno
MNT	Modelo Numérico do Terreno
MODIS	<i>Moderate Resolution Imaging Spectroradiometer</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NSN	Divisor da escala nominal
PND	Fração com dados não uniformes
RCC	Reuniões de classificação e correlação de solos
RedeMDS	Rede Brasileira de Pesquisa em Mapeamento Digital de Solos
RMS	Erro médio quadrado
SAA	Secretaria de Agricultura e Abastecimento

SAD69	<i>South American Datum 1969</i>
SiBCS	Sistema Brasileiro de Classificação de Solos
SIG	Sistema de Informação Geográfica
SIRGAS2000	Sistema de Referência Geocêntrico para as Américas
SRTM	<i>Shuttle Radar Topography Mission</i>
TIN	<i>Triangulated Irregular Network</i>
UM	Unidade de Mapeamento
UTM	Universal Transversa de Mercator

1. INTRODUÇÃO GERAL

O solo tem sido fundamental para a sobrevivência e o progresso da humanidade desde que nossos ancestrais cultivaram a primeira planta há alguns milênios. Hoje ele fornece alimento e outros recursos para mais de 7 bilhões de seres humanos e muitas outras espécies animais, é o meio sobre o qual crescem as plantas e onde também vivem incontáveis outros organismos, e todos o modificam constantemente. Muitas vezes o uso inadequado leva à degradação e compromete sua capacidade de continuar sustentando a nossa e todas as outras formas de vida que dele dependem.

A consciência de que se trata de um recurso finito frente à pressão da crescente população humana, aliada ao reconhecimento de diversas funções ecossistêmicas e inclusive no clima global, recolocaram o solo no foco das questões de sustentabilidade. Nesse sentido, diversas iniciativas têm promovido a percepção da importância do solo pela sociedade. Recentemente, a Assembleia Geral das Nações Unidas aprovou a designação de 2015 como o Ano Internacional do Solo, a partir de uma proposição da FAO e da União Internacional de Ciência do Solo, e 5 de dezembro foi escolhido como o Dia Mundial do Solo. Ações também vêm sendo desenvolvidas no Brasil, sob os auspícios da Sociedade Brasileira de Ciência do Solo (SBCS) em cooperação com várias instituições.

A ênfase na importância da gestão sustentável do solo como base para a produção de alimentos e funções essenciais dos ecossistemas trouxe nova importância à Ciência do Solo. No contexto de gestão sustentável, o conhecimento profundo e amplamente disseminado sobre os solos é indispensável para orientar a definição de estratégias e a tomada de decisões voltadas ao seu uso. Isso demanda a intensificação da pesquisa em diversas áreas e a criação de meios eficazes para disponibilizar os resultados, de modo que possam ser efetivamente utilizados para tal fim. Em 2013, o tema do 34º Congresso Brasileiro de Ciência do Solo, “Ciência do Solo: para que e para

quem?”, refletiu essa preocupação sob a perspectiva da comunidade de cientistas do solo.

Uma das necessidades atuais mais prementes refere-se ao conhecimento amplo e detalhado da distribuição espacial dos solos e de suas propriedades. Por limitações de custo, de tempo e de pessoal, as técnicas convencionais de levantamento mostram-se inviáveis para atender as crescentes demandas por tais informações, na velocidade e com o detalhamento necessário. A pedometria e o mapeamento digital de solos (MDS) surgiram como alternativa para gerar novos mapas de solos, beneficiando-se da disponibilidade cada vez maior de dados espaciais sobre a superfície terrestre e de novos métodos de aquisição em campo, em combinação com ferramentas de processamento avançadas.

Em 2006 o Brasil sediou o segundo *workshop* internacional em MDS, no Rio de Janeiro. Na ocasião foi lançada a proposta de elaborar um novo mapa global retratando algumas propriedades funcionais do solo, culminando na criação do consórcio *GlobalSoilMap.net* no final do mesmo ano. Os estudos apresentados no *workshop* foram publicados em 2008, em um livro com o sugestivo título *Digital Soil Mapping with Limited Data* (Hartemink et al., 2008). Mais recentemente a SBCS criou a Comissão Especializada em Pedometria, dentro da divisão Solo no Espaço e no Tempo, e a EMBRAPA Solos estabeleceu a Rede Brasileira de Pesquisa em Mapeamento Digital de Solos (RedeMDS) visando integrar pesquisadores da área.

Nos últimos anos houve avanços consideráveis em métodos e modelos de predição em MDS, mas ainda persiste a dificuldade em obter informações sobre classes ou propriedades de solos para sua calibração e validação. Uma opção tem sido o uso de informações já disponíveis sobre solos, obtidas por levantamentos convencionais prévios, denominadas de “dados legados”. Entretanto, ela também enfrenta desafios, pois a cobertura e detalhamento são bastante heterogêneos, o vasto conhecimento sobre as relações solo-paisagem e sobre as características e propriedades dos solos que eles sintetizam está em grande parte implícito, e a real qualidade dos mapas convencionais de solos não é conhecida.

A predição de classes ou propriedades a partir de mapas de solos existentes depende da recuperação do modelo mental de solo-paisagem usado

na sua criação. Muitas vezes esse processo é conduzido sem considerar as particularidades e limitações dos dados existentes, o que pode comprometer os resultados das predições. Com vistas a poder tirar o máximo proveito de levantamentos existentes, ainda são necessários estudos para verificar o conhecimento que pode ser resgatado dos mesmos, bem como até que ponto as técnicas de MDS possibilitam enriquecê-los ou gerar novos dados de forma confiável, levando em conta a qualidade original.

1.1. Objetivo geral

O objetivo geral desta pesquisa foi aplicar, medir e avaliar metodologias e ferramentas de mapeamento digital de solos para a predição de classes de solo a partir de levantamentos convencionais em diferentes escalas na região da Serra Gaúcha, Rio Grande do Sul. A premissa foi que o conhecimento pedológico embestado em mapas convencionais de solos, produto de um modelo mental solo-paisagem desenvolvido pelos autores durante o processo de mapeamento, pode ser resgatado, sistematizado e reutilizado para fazer novas predições com base em covariáveis relativas aos principais fatores de formação dos solos nas áreas em questão.

1.2. Estrutura da tese

Esta tese de doutorado está organizada em capítulos na forma de artigos com vistas a facilitar a divisão dos assuntos e a publicação dos resultados. O Capítulo I contém uma revisão bibliográfica reunindo os principais conceitos e paradigmas da ciência do solo, desde seu surgimento até os avanços recentes em mapeamento digital de solos (MDS). Ele também enfatiza a importância dos levantamentos existentes como valiosos dados de entrada para ajustar modelos de predição em MDS a fim de suprir as demandas por informações sobre os solos, as principais estratégias de uso e as formas de avaliar a qualidade dos resultados das predições.

O Capítulo II trata da caracterização e avaliação de mapas legados de solos, cujos resultados podem ser críticos para o uso desses mapas em diferentes aplicações, especialmente no desenvolvimento de modelos de predição em MDS. São propostos, analisados e discutidos indicadores calculados a partir dos polígonos e da legenda dos mapas, alguns dos quais podem ser de grande utilidade para orientar o uso consciente de mapas de solos existentes. O artigo correspondente a este capítulo foi publicado no

volume 38 da Revista Brasileira de Ciência do Solo, em dezembro de 2014, previamente à apresentação e defesa desta tese.

O Capítulo III corresponde a um estudo voltado ao aperfeiçoamento de mapas legados de solos por meio de um procedimento de desagregação de unidades de mapeamento combinadas, buscando individualizar as classes que as compõem a fim de produzir um mapa de solos mais detalhado que o original. A abordagem proposta é baseada no resgate criterioso do conhecimento local das relações solo-paisagem combinado com técnicas de MDS, usando como área de estudo um mapa de um levantamento semidetalhado avaliado no Capítulo II. O procedimento adotado, guiado por conhecimento formal, possibilita refinar sensivelmente mapas de solos existentes de forma a agregar valor para diversos usos futuros, tal como demonstrado em um exemplo de avaliação de aptidão para a viticultura.

O Capítulo IV aborda a predição de classes de solos em áreas não visitadas a partir de áreas de referência onde há levantamentos convencionais de alta intensidade, em um processo análogo à inferência efetuada por pedólogos a partir de pontos visitados no terreno. A abordagem utilizada baseia-se em mineração de dados para resgatar automaticamente o conhecimento sobre as relações solo-paisagem e reapplicá-lo em áreas vizinhas com relevo semelhante, usando um mapa de um levantamento detalhado também avaliado no Capítulo II. Além da viabilidade de extrapolar regras solo-paisagem com alta resolução espacial, é investigado o impacto da densidade e do esquema de amostragem, do número e combinação de covariáveis preditoras e da resolução espacial sobre a qualidade das predições. Adicionalmente, por meio de múltiplas realizações, são computados mapas para descrever a variação espacial da confiança das classes de solo preditas.

Por fim, são elaboradas conclusões gerais e feitas algumas considerações sobre estudos futuros, com base nos resultados obtidos nos três capítulos anteriores, e em seguida é apresentada a relação de referências bibliográficas utilizadas.

2. CAPÍTULO I – O SOLO: DA OBSERVAÇÃO EMPÍRICA AO MAPEAMENTO DIGITAL

2.1. O estudo do solo como ciência

O solo tem sido objeto de estudo há muitos séculos, embora durante a maior parte do tempo não houvesse uma ciência formal dedicada a esse propósito. Os gregos provavelmente tenham sido os primeiros a reconhecer que informações úteis poderiam ser obtidas da observação de uma seção vertical do solo, ou seja, do perfil. Theophrastos (371-286 a.C.), um aluno de Aristóteles, deu ao solo o nome “edaphos” e supôs uma estratificação no seu conteúdo de húmus, um subsolo que supria nutrientes à grama e às ervas, um substrato que fornecia nutrientes às árvores e, abaixo delas, o desconhecido. Além de descrever várias propriedades dos solos e suas relações com as plantas, ele definiu seis grupos de terras aptas para diferentes tipos de cultivos e avaliou quantitativamente seu potencial produtivo. Em Roma, Cato (234-149 a.C.) desenvolveu um sistema funcional de classificação de solos aráveis com nove classes e 21 subclasses. Varro (116–27 a.C.) dedicou-se à composição física dos solos, Columella (4-70) estudou as propriedades físicas e Plínio (23-79) dedicou muita atenção às rochas e minerais como origem da formação dos solos. Vários estudiosos romanos empiricamente derivaram relações que orientaram o uso e a conservação da terra para a produção de alimentos e fibras por longo tempo. O conhecimento antigo dos solos era extenso e uma terminologia pedológica já estava estabelecida. Com a queda de Roma, no entanto, o estudo do solo agrícola estagnou e poucas contribuições foram feitas até a metade do século XIX, quando algumas classificações utilitárias foram novamente propostas e Justus von Liebig (1803-1873) publicou seus primeiros estudos sobre química do solo e nutrição mineral de plantas (Arnold, 1983; Buol et al., 2011; Miller & Schaetzl, 2014).

O estudo do solo como uma ciência formal, a pedologia, começou somente no final do século XIX, e seu início é atribuído ao cientista russo Vasili

Vasilevich Dokuchaev (1846–1903). Em 1883 Dokuchaev publicou um estudo dos solos escuros das estepes russas, os Chernozem, no qual ele desenvolveu e usou conceitos sobre a natureza e a gênese de perfis de solo e sua relação com a paisagem. Junto com seus estudantes, ele aplicou princípios de morfologia para descrever os principais grupos de solos, elaborou a primeira classificação científica e desenvolveu métodos de mapeamento, estabelecendo os fundamentos para a moderna ciência do solo. Dokuchaev é reconhecido por desenvolver a nomenclatura básica dos horizontes A, B e C, e um modelo fatorial de desenvolvimento no qual os solos e seus padrões eram vistos como resultado da interação de vários fatores do ambiente. Em 1886 ele propôs que “solo” deveria ser usado como um termo científico, e mais tarde postulou que os solos não eram simplesmente um amontoado de materiais geológicos, mas objetos únicos produzidos pela ação combinada da atividade de animais, plantas, clima, relevo e geologia ao longo do tempo (Arnold, 1983; Phillips, 1998; Buol et al., 2011).

Dokuchaev reconheceu explicitamente a importância de vários cientistas no desenvolvimento de suas ideias, mas ele foi influenciado de forma profunda por um número muito maior deles. A concepção do solo como um corpo natural em evolução na paisagem surgiu no vácuo da impactante teoria darwinista de evolução das espécies, publicada pela primeira vez em 1859 (Espíndola, 2010). Charles Darwin (1809-1882) talvez seja o mais subestimado cientista do solo. Com grande domínio da geologia em função de seus estudos com fósseis, ele tinha o conceito de horizontes e propunha que o solo era produto de transformações biomecânicas das rochas ao longo do tempo. Os registros de uma de suas escavações em 1881 já mostravam uma sucessão de camadas A, B, C e D, bem como suas respectivas constituições. No mesmo ano Darwin publicou um estudo mostrando como a ação das minhocas modificava a textura, dando ênfase à origem biológica de várias propriedades do solo em escala local. O desinteresse quase total pelo seu trabalho fez com que o papel dos fatores abióticos fosse superestimado durante décadas, e apenas recentemente o papel crucial da biota foi retomado com mais ênfase na ciência do solo (Schaetzl & Anderson, 2005; Johnson & Schaetzl, 2015).

Alexander von Humboldt (1769-1859) influenciou Dokuchaev ainda mais diretamente com seus estudos sobre clima e vegetação, após uma

viagem de cinco anos pela América do Sul e México. Humboldt fez medições meteorológicas, quantificou correlações entre clima e formas vegetativas e usou o conhecimento obtido desses estudos para delinear zonas de clima-vegetação, dando origem à biogeografia. Em 1817, Humboldt publicou um artigo sobre isotermas que impulsionou o uso de isolinhas para representação geográfica e análise de inúmeros outros fenômenos. Em 1829 ele foi convidado a visitar as minas da Sibéria e dos montes Urais e a apresentar algumas conferências pela Rússia, terminando por convencer o governo russo a implantar uma rede de estações meteorológicas. Dokuchaev empregou a mesma técnica de isolinhas no seu mapa de isohúmus para ilustrar os resultados de seu estudo pioneiro sobre os solos Chernozem, publicado em 1883. Com recursos limitados e somente alguns percursos na vasta região que estudou, Dokuchaev conseguiu construir o primeiro mapa coerente de húmus dos Chernozem baseado no princípio de correlação climática de Humboldt, predizendo locais desconhecidos com base na disponibilidade de dados climáticos e em relações solo-vegetação-clima. A rede de estações meteorológicas que Humboldt propôs em 1829 forneceu os dados necessários (Brown, 2006; Hartemink et al., 2013).

O desenvolvimento da ciência do solo subsequente a Dokuchaev teve a contribuição de enorme contingente de cientistas e acabou por definir várias áreas de pesquisa, destacando-se a gênese, a morfologia e classificação, e o mapeamento dos solos. Embora elas sejam inseparáveis por se referirem a diferentes aspectos do mesmo corpo natural, cada uma tem e continua tendo sua própria história, conceitos e desafios. Sustentando sua base teórica estão as áreas fundamentais da física, química, mineralogia e biologia do solo.

2.1.1. Gênese dos solos

Conceitos sobre a formação dos solos similares aos de Dokuchaev, embora menos abrangentes, haviam sido desenvolvidos aproximadamente na mesma época por cientistas em outros países. Com a difusão restrita pela dificuldade do idioma russo, a importância das ideias de Dokuchaev passou a ser mais amplamente reconhecida apenas a partir de 1914, quando seu aluno Konstantin Glinka (1863–1935) publicou uma versão revisada dos estudos em alemão. Estimulado pelo conteúdo do livro de Glinka, o então líder do *US*

Bureau of Soils, Curtis Marbut (1863–1935), resolveu traduzi-lo para o inglês. Sua publicação em 1927 tornou os conceitos de Dokuchaev acessíveis a um grande número de cientistas do solo, por quem eles foram muito bem recebidos (Kellog, 1974).

A idéia de solo como produto de fatores do ambiente instigou diversas propostas de formulações matemáticas, e o próprio Dokuchaev propôs uma equação em 1899. Outros cientistas russos sugeriram algumas variantes depois dele e, no congresso internacional de ciência do solo de 1932, Chas F. Shaw (1881-1939) apresentou ao público uma proposta de formulação matemática para os fatores de formação, representada na equação 1:

$$S = m(c + v)t + d \quad 1$$

onde S é o solo, m o material de origem, c o clima, v a vida orgânica, t o tempo e d os processos de erosão e deposição. Nessa época, se considerava que os fatores de formação eram as causas da formação dos solos e as propriedades destes eram seus efeitos, e que os primeiros definiam as condições para os processos internos de formação dos solos. Em 1941, Hans Jenny (1899–1992) concluiu que os fatores de formação não são forças ou causas, mas variáveis independentes, e reescreveu a função de Shaw 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 2:

$$S = f(cl, o, r, p, t...) \quad 2$$

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. As reticências no final da equação denotam que o solo é um sistema aberto, cuja formação pode eventualmente incluir outros fatores, tais como tempestades de poeira e erupções vulcânicas. 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, significa que as relações entre eles podem ser mensuradas e quantificadas (Hudson, 1992; Scull et al., 2003; Buol et al., 2011).

A noção de fatores de formação existia há bastante tempo, mas com diferentes concepções sobre como eles se inter-relacionam para determinar a formação dos solos. Fundamentada nas ideias iniciais de Dokuchaev, a

equação de Jenny foi a formulação mais bem sucedida e propiciou uma base para interpretar os atributos do solo em termos de sua história genética. Ela abriu caminho para o estudo e a compreensão dos processos de formação dos solos (processos pedogenéticos) que levam à diferenciação dos horizontes, e permanece como o principal modelo explanatório dos solos até os dias de hoje (Bockheim et al., 2005).

2.1.2. Morfologia e classificação

Assim como ocorreu em outras ciências naturais, à medida que o estudo e compreensão sobre a formação do solo avançaram, surgiu a necessidade de descrever, diferenciar e classificar esse objeto de estudo. Houve muitos sistemas de classificação dos solos antes de Dokuchaev, em geral classificações técnicas ou utilitárias voltadas ao seu uso. Em meados do século XIX, vários países da Europa haviam efetuado levantamentos e desenvolvido classificações com viés agroecológico. A classificação natural ou científica do solo, por outro lado, busca a categorização de solos em grupos em diferentes níveis de generalização de acordo com suas propriedades intrínsecas. Seus objetivos são organizar o conhecimento, facilitar a leitura e a interpretação de propriedades, a clareza na compreensão de relacionamentos e a comunicação e a transferência de tecnologia (Buol et al., 2011).

Da mesma forma que a classificação de plantas e animais, a morfologia se tornou a base mais óbvia para desenvolver sistemas de classificação visando agrupar os diferentes tipos de solos de acordo com suas propriedades. A morfologia do solo abrange a cor, a estrutura física, as propriedades químicas e mineralógicas do material do solo, a associação espacial dos materiais em horizontes e a dinâmica da temperatura e umidade do solo *in situ* (Buol et al., 2011; Legros, 2012). Dokuchaev havia desenvolvido uma classificação útil com base na morfologia, mas nela os horizontes do perfil eram considerados principalmente com propósito descritivo.

Enquanto traduzia o livro de Glinka, em 1921, Marbut foi o primeiro a sugerir um sistema em que os horizontes fossem usados para classificar e diferenciar os solos pelas suas próprias características (Kellog, 1974). Apesar de entender o solo como um corpo natural em evolução, Marbut concebia solos “maduros”, ou solos “modais”, que deveriam ser classificados estritamente com base nas características dos perfis. Sem fazer referência à gênese, ele

desconsiderava que os solos podem mudar com o tempo, e que essa mudança pode ser drástica sob a ação humana (Schaetzl & Anderson, 2005). Um exemplo contemporâneo é dado por Streck et al. (2008) na Encosta Inferior do Nordeste do Rio Grande do Sul, onde a erosão do horizonte A chernozêmico de Chernossolos Argilúvicos causada pelo uso agrícola os transformou em Cambissolos Hápicos ou em Luvisolos Hápicos em algumas décadas.

O sistema proposto por Marbut em 1935, mesmo conflitante com a ideia dos fatores de formação, foi o primeiro passo para a posterior constatação de que perfis de solo constituem um registro da gênese e história do solo, possivelmente uma das descobertas mais significativas na pedologia. A partir de então, os horizontes se tornaram camadas genéticas que poderiam ser (1) distinguidas com base em propriedades morfológicas, químicas e físicas e (2) usadas para interpretar a história evolutiva do solo. Dessa forma, os horizontes se tornaram efetivamente a base para a estruturação dos sistemas de classificação taxonômica de solos no mundo todo (Bockheim et al., 2005; Olson, 2006).

No entanto, a classificação dos solos é um desafio prático porque eles não existem na natureza como entidades discretas ou individuais, tais como plantas ou animais, e sim como corpos tridimensionais contínuos. A necessidade de fazer seu reconhecimento e identificação em campo para fins de classificação levou ao desenvolvimento do conceito da menor unidade tridimensional reconhecível que pode ser classificada, o *pedon*, que veio a se tornar a unidade básica de amostragem. Um agrupamento de *pedons* semelhantes define uma área denominada *polipedon*, um corpo de solo cujos limites laterais coincidem com os de outros *polipedons* e cuja profundidade é determinada pelos *pedons* que o constituem (Ahrens & Arnold, 2012; Legros, 2012).

Ainda antes do *pedon*, em 1935, Geoffrey Milne (1898–1942) introduziu o conceito de catena para descrever a variação de padrões de profundidade e drenagem do solo observadas ao longo de uma encosta, do topo até a base, ou seja, uma topossequência. Esse conceito ressaltou a importância local da topografia no desenvolvimento dos solos, já observada por Jenny, e propiciou aos pedólogos uma base para estimar sua ocorrência a partir de padrões repetidos no relevo, tornando-se o primeiro modelo efetivo de

solo-paisagem. Mais tarde, a constatação de que certos processos também se propagam lateralmente culminou na definição do conceito de superfície geomórfica: uma área identificável no terreno, resultante de um conjunto comum de processos erosionais e deposicionais ao longo do tempo. As superfícies geomórficas possibilitaram compreender, interpretar e transferir o conhecimento adquirido em uma área a outras semelhantes, tornando-se um modelo solo-paisagem universal para orientar a identificação, classificação e delineamento dos solos (Buol et al., 2011; Wysocki et al., 2012; Bockheim et al., 2005).

Um marco no estabelecimento de sistemas consistentes de classificação taxonômica dos solos foi a publicação do *Soil Taxonomy* pelo Departamento de Agricultura dos EUA em 1975, que se tornou amplamente usado no mundo. Ele emprega critérios morfológicos quantitativos para definir tipos de solo e também usa conceitos de gênese do solo para guiar a seleção e aplicação ordenada desses critérios (Buol et al., 2011). No Brasil, as primeiras tentativas de estruturação de um sistema de classificação de solos iniciaram logo após a criação da Comissão de Solos do Centro Nacional de Ensino e Pesquisas Agronômicas do Ministério da Agricultura, em 1947. Várias adequações foram depois introduzidas com base em observações feitas durante os levantamentos de reconhecimento de solos dos estados Rio de Janeiro e de São Paulo, publicados respectivamente em 1958 e 1960. A organização do Sistema Brasileiro de Classificação de Solos (SiBCS) progrediu na década de 1970, quando os levantamentos de solos no país tomaram maior impulso, sendo a 1ª aproximação publicada em 1980. A classificação dos solos vigente no país absorve muitos dos conceitos e critérios do *Soil Taxonomy* (Oliveira, 2001; IBGE, 2007; Embrapa, 2013). Nos seus fundamentos, os dois sistemas de classificação são genéticos, pois as classes refletem a predominância de certos processos pedogenéticos (Arnold, 1983; Schaetzl & Anderson, 2005; Legros, 2012).

A taxonomia não é fechada, ela está em constante evolução e aprimoramento para acomodar novas descobertas e hipóteses ou para contemplar necessidades de diferenciação ou de reagrupamento de certos solos. O *Soil Taxonomy*, por exemplo, teve uma segunda edição revisada publicada em 1999, e as chaves de classificação são revisadas e publicadas

periodicamente (Buol et al., 2011). No Brasil, desde 1995 o SiBCS também é revisado e atualizado, contando com a colaboração de um grande número de cientistas do solo de várias instituições de ensino e pesquisa no país. Necessidades ou sugestões de aprimoramento são identificadas nas reuniões de classificação e correlação de solos (RCC) ou no blog do SiBCS, avaliadas pelo Comitê Executivo e, quando há consenso, muitas são incorporadas. Atualmente o SiBCS encontra-se na terceira edição (Embrapa, 2013).

Qualquer sistema de classificação é um reflexo da situação do conhecimento existente sobre os objetos classificados em dado momento, estruturado de uma maneira sistemática para facilitar a comunicação com outros cientistas e, mais importante ainda, com estudantes em vias de se tornar futuros cientistas. O fato de que sistemas de classificação de solos mudam com o tempo é um resultado inevitável da evolução da pesquisa na ciência do solo e em outras áreas relacionadas a ela e ao uso do próprio solo. O desenvolvimento de novos métodos analíticos para medir a composição física, química e mineralógica dos horizontes também impulsiona a revisão dos sistemas, pois oferece novas possibilidades de distinguir suas propriedades (Buol et al., 2011; Embrapa, 2013; Hartemink & Minasny, 2014).

2.1.3. Mapeamento

Desde o início da agricultura, as populações desenvolviam mapas mentais para definir locais apropriados para plantar e se estabelecer, a partir de observações do ambiente. Por volta de 500 a.C. os gregos já descreviam relações espaciais entre a vegetação, o clima e os tipos de solo, e observavam que alguns eram mais apropriados a certas culturas que outros. Eles tinham os conceitos de associação espacial e funcional e os princípios de projeções, mais tarde consolidados por Cláudio Ptolomeu (~100-170), mas preferiam registrar a ocorrência dos solos de forma descritiva porque não havia como elaborar mapas confiáveis. Vários outros povos também tentaram elaborar mapas de solos, em geral retratando a qualidade das terras para a produção agrícola e em áreas de abrangência limitada. Muitas vezes as parcelas de terra cultivada eram usadas como unidade de mapeamento. Alguns mapas abrangeram áreas maiores, mas suas escalas eram reduzidas e eles continham distorções severas na representação (Arnold, 1983; Hartemink et al., 2013; Miller & Schaetzl, 2014).

Em um contexto científico, o mapeamento efetivo de certo objeto ou fenômeno envolve a determinação da posição geográfica de seus limites, o que só foi possível com razoável grau de acurácia a partir do século XVIII, mas com métodos ainda extremamente trabalhosos e caros. Foi somente no século XIX, com o advento de mapas topográficos de maior precisão, que o mapeamento de solos passou a dispor de um suporte cartográfico confiável. Esses mapas continham elementos que podiam servir de referência para localizar as feições observadas em campo, tais como cursos d'água, estradas, curvas de nível e cidades, facilitando assim o delineamento dos solos. O famoso mapa de isohúmus publicado por Dokuchaev em 1883 só pôde ser elaborado graças à disponibilidade de uma base cartográfica adequada na área que ele estudou, sobre a qual ele pôde traçar as isolinhas de forma consistente (Schaetzl & Anderson, 2005; Miller & Schaetzl, 2014).

Com o desenvolvimento da pedologia desde então, o mapeamento de solos tem buscado delinear a ocorrência de tipos de solos no ambiente com vistas a efetuar um registro cartográfico da sua distribuição espacial. Essa tarefa demanda o domínio dos conceitos da gênese e da morfologia e classificação, para identificar os solos e compreender sua associação com o ambiente, e necessita de um suporte para localizar espacialmente seus limites presumidos. O paradigma solo-paisagem propiciou uma base consistente para prever o arranjo dos solos a partir de características da paisagem mais facilmente observáveis, e tem sido a estratégia dominante para discriminar unidades de mapeamento de solos por quase um século. O suporte espacial para materializar os limites das unidades de mapeamento, por outro lado, mudou consideravelmente ao longo do tempo desde Dokuchaev, e continua evoluindo rapidamente (Hudson, 1992; Bui, 2004, Schaetzl & Anderson, 2005).

A disseminação das fotografias aéreas representou um salto qualitativo no mapeamento de solos, em comparação com o uso de mapas-base. Após a primeira guerra mundial, na qual elas haviam sido intensamente empregadas, vários países passaram a utilizá-las para mapeamentos em geral. Um estudo nos EUA em 1929 mostrou que o uso de fotografias aéreas para apoiar o mapeamento de solos incrementava a eficiência, a qualidade e a consistência do mapa final. Além disso, seu uso reduzia a necessidade de bases cartográficas pré-existentes, o que representava economia considerável.

Em 1932, Michigan tornou-se o primeiro estado norte-americano a relatar o uso de um estereoscópio de espelhos para delimitar os solos em áreas de vegetação densa e de difícil acesso. As imagens tridimensionais estereoscópicas facilitaram a discriminação das superfícies geomórficas, possibilitando reduzir a densidade de observações em campo para inferir a distribuição dos solos. Finalmente, em 1951, o *U.S. Soil Survey Manual* passou a incluir padrões e recomendações para a interpretação de fotografias aéreas, e desde então elas se tornaram a base dominante para o delineamento dos solos em todo o mundo (Miller & Schaetzl, 2014).

A introdução recente de tecnologias de informação geográfica, tais como o Sistema de Posicionamento Global (GPS), os Sistemas de Informação Geográfica (SIG) e dispositivos de sensoriamento remoto, está instigando outra revolução no mapeamento. Assim como ocorreu com as fotografias aéreas e os mapas topográficos antes delas, essas tecnologias têm aberto novas possibilidades no mapeamento dos solos, fornecendo um vasto conjunto de dados espaciais e novos métodos de processamento e de análise desses dados. Elas demandam a reavaliação de alguns procedimentos atuais e o desenvolvimento de novas técnicas para o avanço da ciência do solo. Os conceitos básicos da pedologia permanecem os mesmos, mas as possibilidades abertas com esses novos recursos estão redefinindo o mapeamento dos solos novamente, levando inclusive ao estabelecimento de uma nova área, o mapeamento digital de solos (Scull et al., 2003; Lagacherie & McBratney, 2007; Miller & Schaetzl, 2014).

2.2. Levantamentos de solos

Um levantamento de solos tem como objetivo principal efetuar um inventário dos solos existentes em determinada área, uma tarefa complexa que requer o domínio simultâneo dos principais conceitos da gênese, da morfologia e classificação dos solos, bem como das relações do solo com o ambiente no qual ele se formou. 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. Trata-se, portanto, de um estudo do ambiente baseado em dois paradigmas complementares: o dos fatores de formação, para reconhecer e explicar os solos, e das relações solo-paisagem, para inferir e descrever sua distribuição espacial. Tipicamente,

seus resultados incluem um mapa de solos e um relatório textual que sintetiza o conhecimento adquirido, contendo a caracterização das classes de solo identificadas, a composição das unidades de mapeamento, a descrição da ocorrência dos solos na paisagem e dados analíticos e descritivos de perfis representativos (Hudson, 1992; Bui, 2004).

Nos Estados Unidos, relatórios de levantamentos de campo e mapeamentos de solos começaram a ser produzidos de forma sistemática pelo Departamento de Agricultura em 1899. Desde o início da década de 1950, o *US National Soil Survey Handbook* e o *Soil Survey Manual* definem padrões e especificações revisados regularmente para orientar a condução de todas as fases de levantamentos de solos pelo *US National Cooperative Soil Survey* (Hempel et al, 2008). No Brasil, as primeiras análises físicas e químicas foram realizadas em 1885, em amostras de terra de São Paulo. A primeira tentativa de análise de um perfil provavelmente ocorreu em 1928, também em São Paulo, e a confecção do primeiro mapa agrológico de solos em 1947, na Paraíba. Nesse mesmo ano foi criada a Comissão de Solos do Centro Nacional de Ensino e Pesquisas Agronômicas do Ministério da Agricultura, quando os levantamentos de solos no país tomaram impulso (Santos, 1997; IBGE, 2007). Desde então, também começaram a ser estabelecidos critérios e especificações para levantamentos em diferentes escalas, os quais são aperfeiçoados continuamente e reunidos em publicações de referência (IBGE, 2007; Embrapa, 2013).

De forma geral, um levantamento de solos inicia pela reunião e estudo dos dados disponíveis (cartas topográficas, fotografias aéreas, outros mapas) para definir unidades básicas de mapeamento, a partir das quais locais são escolhidos e visitados no campo para observar e descrever perfis representativos. 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 delinear as unidades de mapeamento definitivas. 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 certo tipo de solo termina e outro começa (Hudson, 1992; Bui, 2004).

A forma de execução praticamente não mudou desde os primeiros levantamentos sistemáticos de solos, apesar do desenvolvimento tecnológico verificado em outras áreas (Miller & Schaetzl, 2014). A necessidade de recursos financeiros, de pessoal capacitado e de tempo restringiu o nível de detalhamento e a expansão dos levantamentos, restando muitas áreas não cobertas e uma escassez generalizada de mapas detalhados. Um declínio acentuado ocorreu a partir de meados da década de 1980, quando os levantamentos de solos praticamente estagnaram (McBratney et al., 2003; Hartemink & McBratney, 2008). Sanchez et al. (2009) estimaram que atualmente cerca de 109 países possuem mapas de solos em escalas 1:1.000.000 ou maiores, mas eles cobrem somente 31% da superfície da Terra emersa e livre de gelo. Hartemink (2008) mostrou que apenas 10% a 30% da área dos países da América do Sul, sudeste da Ásia e da África está contemplada com escalas maiores que 1:100.000. O Brasil está integralmente mapeado nas escalas 1:5.000.000 e 1:1.000.000, mas apenas 35% do território está coberto com escalas entre 1:100.000 e 1:600.000, e uma porção muito restrita tem mapas de solos mais detalhados (Mendonça-Santos & Santos, 2007).

Mesmo com lacunas na cobertura e frequentemente com pouco detalhamento, os mapas e relatórios produzidos por levantamentos convencionais de solos ainda são a maior fonte de informação sobre esse recurso natural complexo e vital. Em muitos casos, eles constituem o único registro da natureza altamente variável das relações solo-paisagem (Scull et al., 2005; Hempel et al., 2008; Yang et al., 2011). Devido à sua importância, a partir da década de 1990 vários países investiram na criação de sistemas de informações de solos, contendo mapas e perfis georreferenciados e tabelas com dados analíticos e descritivos. Mundialmente a FAO (*Food and Agriculture Organization*) desenvolve iniciativa semelhante, e no Brasil a Embrapa vem estruturando um banco de dados nacional de perfis de solos (Lagacherie & McBratney, 2007; Mendonça-Santos & Santos, 2007; Omutu et al., 2013).

Embora tenham possibilitado avanços, tais iniciativas também evidenciaram limitações de levantamentos convencionais em suprir a crescente

e diversificada demanda por informações sobre os solos, tanto por questões de custo e tempo de execução quanto pelas características dos produtos (McBratney et al., 2012; Omuto et al., 2013). Mapas convencionais representam a distribuição complexa e contínua dos solos na paisagem na forma de polígonos discretos, que delimitam uma ou mais classes de solo. A localização das classes em unidades de mapeamento combinadas não é conhecida, a não ser de forma descritiva. A escala de publicação costuma ser determinada com base no material cartográfico disponível, não em normas cartográficas estritas e frequentemente há inconsistências entre mapas adjacentes, mesmo com a mesma escala de publicação. Diferenças no detalhamento das informações descritivas nos relatórios de levantamento também são comuns, e em geral não há uma avaliação quantitativa da acurácia espacial e temática do mapa (Heuvelink & Webster, 2001; Shi et al., 2004; Bui, 2004; Hengl & Husnjak, 2006; Hempel et al., 2008; Nussbaum et al., 2011; Miller, 2012; Phillips, 2013; Sarmiento et al., 2014).

A variedade de aplicações analíticas em modelagem que emergiu nas últimas duas décadas requer informações espacialmente mais detalhadas e específicas sobre o recurso solo. Ela abrange, além da produção de alimentos e fibras, a conservação de habitats e do próprio solo, recursos hídricos, avaliações ambientais, mudanças climáticas, bioenergia e diversas outras áreas. A maioria das aplicações necessita de mapas que descrevam a variação espacial contínua de determinadas propriedades dos solos para alimentar seus modelos, mas mapas de classes de solo individualizadas com maior resolução espacial, assim como dados analíticos de perfis de solos georreferenciados, também constituem parcela expressiva das demandas. A disponibilidade atual de tecnologias para coleta e análise de dados espaciais, tais como SIG, GPS e dispositivos de sensoriamento remoto, tornam o mapeamento digital de solos a alternativa potencialmente mais eficaz para tentar suprir essa diversidade de novas demandas (McBratney et al., 2003; McBratney et al., 2012; Omuto et al., 2013).

2.3. Mapeamento digital de solos

Conforme 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 sobre os solos e de covariáveis ambientais disponíveis. Em uma definição mais sintética, ele se refere à geração de bases espaciais de solos por meio de relações estatísticas quantitativas entre observações sobre o solo e outros dados ambientais (McBratney et al., 2012).

O mapeamento digital de solos representa um novo estágio na evolução dos procedimentos e metodologias da ciência do solo, com base no conhecimento acumulado e nas tecnologias disponíveis. Os paradigmas dos fatores de formação e das relações solo-paisagem permanecem os mesmos, a principal diferença é que no lugar de avaliações mentais qualitativas busca-se empregar modelos quantitativos de base matemática. Na essência, as principais etapas na rotina de mapeamentos digitais são análogas àquelas usadas na elaboração de mapas de solos em levantamentos convencionais. Assim como nestes, o mapeamento digital faz uso de informações sobre os solos coletadas ou disponíveis em alguns locais. Essas informações são utilizadas para ajustar um modelo quantitativo com base em covariáveis relativas às condições do ambiente nos mesmos locais, e o modelo ajustado é depois empregado para prever propriedades ou classes de solos no restante da área de interesse (Grunwald, 2006; Lagacherie, 2008; MacMillan, 2008).

As primeiras proposições surgiram na década de 1970, sob a denominação de mapeamento preditivo de solos, e progrediram nos anos 1980 com a disseminação dos SIG, do GPS e das imagens de sensoriamento remoto orbital. Um grande impulso ocorreu a partir da década de 1990, com os avanços na tecnologia de computação, a implantação de sistemas de informação de solos e a evolução de programas de estatística e modelagem. Posteriormente, a rede mundial de computadores agilizou o intercâmbio e a busca de consensos e padrões operacionais, e atualmente inovações vem ocorrendo na aquisição de dados (McBratney et al., 2003; Scull et al., 2003; Bockheim et al., 2005; Rossiter, 2005; Sanchez et al., 2009, Omuto et al., 2013; Hartemink & Minasny, 2014). Os progressos tecnológicos e em modelagem resultaram no advento da pedometria como ramo da ciência do solo voltado à aplicação de métodos matemáticos e estatísticos para o estudo da gênese e distribuição dos solos. Além de relações entre covariáveis predictoras e classes ou propriedades, ela também inclui novos métodos e técnicas para a aquisição

de informações sobre o solo *in situ* (McBratney et al., 2000; Hartemink & Minasny, 2014).

Uma das abordagens conceituais mais difundidas para mapeamento digital de solos é o modelo *s.c.o.r.p.a.n* (McBratney et al., 2003), representado na equação 3:

$$S_{c,p} = f(s.c.o.r.p.a.n) \quad 3$$

onde $S_{c,p}$ se refere a uma classe solo ou a uma propriedade do solo a ser predita, s é a classe ou propriedade conhecida em um dado ponto, c é o clima (variáveis climáticas do ambiente em um dado ponto), o se refere aos organismos (incluindo cobertura vegetal e uso do solo), r é o relevo (atributos do terreno e/ou classes de formas), p é o material de origem (litologia), a é uma variável referente à idade do solo (tempo) e n se refere à localização espacial (georreferenciamento).

Adaptado da equação clássica dos fatores de formação 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. Ele expande a equação de Jenny a fim de incorporar informações sobre classes ou propriedades dos solos observadas em locais específicos e de incluir explicitamente a posição espacial (n). Essa não é uma ideia nova, pois o próprio Jenny chegou a mensurar e validar relações espaciais (Hudson, 1992), mas é uma nova formulação para a importância dessas relações e seu potencial para explicar as propriedades do solo. O modelo *s.c.o.r.p.a.n* reflete a ênfase na localização espacial como um atributo chave capaz de acrescentar poder preditivo nos modelos, com base em um vasto conjunto de geotecnologias.

2.3.1. Fatores *s.c.o.r.p.a.n* e atributos do terreno

O pressuposto do modelo *s.c.o.r.p.a.n*, como qualquer modelo preditivo, é que as covariáveis preditoras possam ser obtidas mais facilmente e a menor custo do que a aquisição da informação a ser predita. Atualmente mapas de covariáveis relativas ao clima (temperatura, precipitação, evaporação, etc.) podem ser obtidos por interpolação a partir de dados de estações meteorológicas de superfície ou derivados das imagens de alguns satélites. Imagens de sensoriamento remoto multiespectral também

representam uma fonte importante de covariáveis sobre os organismos (vários índices de vegetação, uso do solo, etc.) e sobre o material de origem (litologia, mineralogia). Mapas geológicos ou geomorfológicos existentes também são úteis para caracterizar o material de origem, bem como a idade do solo indiretamente. O georreferenciamento dos dados garante a consideração da posição espacial, e o relevo pode ser descrito por inúmeras covariáveis computadas a partir de uma gama de dados altimétricos disponíveis ou facilmente adquiríveis (Hansen et al., 2009; Sanchez et al., 2009; Böttinger et al., 2010; Stum et al., 2010; Nauman et al., 2014, Odgers et al., 2014a).

Apesar da disponibilidade de dados para caracterizar vários fatores de formação, o relevo continua sendo a base para a maioria dos modelos de predição de classes ou de propriedades do solo. Isso reflete a manutenção do paradigma solo-paisagem, que por décadas tem sido a base para identificar e descrever a ocorrência e distribuição espacial dos solos. O termo paisagem é usualmente empregado para descrever o relevo (ou topografia) através da síntese de algumas características do terreno, como posição relativa na encosta e forma da superfície. Reconhecido como fator pedogenético dominante em escala local, o relevo 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 (Moore et al., 1991; Moore et al., 1993; MacMillan et al., 2008; Wysocki et al., 2012).

Em meio digital o relevo é representado por modelos numéricos contendo uma estimativa da altitude em cada local e que recebem várias denominações, tais como modelos digitais de elevação (MDE), modelos digitais do terreno (MDT) e modelos numéricos do terreno (MNT), entre outros. Existem várias estruturas para armazenamento e representação digital do relevo, sendo mais utilizada a estrutura *raster* (matricial) pela simplicidade e facilidade de processamento. Nessa forma de armazenamento, também adotada para outros dados espaciais como imagens de sensoriamento remoto, o valor da altitude é armazenado na forma de uma matriz com células (*pixels*) quadradas (Moore et al., 1993; Burrough & McDonnell, 1998).

Além de representar a variação espacial da altitude, a modelagem digital do terreno possibilita derivar quantitativamente uma considerável variedade de atributos topográficos e hidrológicos do terreno. Essa é uma área

em franco desenvolvimento e de grande interesse da pedometria, pois atributos do terreno podem ser usados não apenas para a predição de classes ou propriedades de solos, mas também como indicadores de diversos processos ambientais e, portanto, para explicar a própria gênese dos solos (Moore et al., 1993; McKenzie & Ryan, 1999; Burrough et al., 2000; MacMillan et al., 2008; Debella-Gilo & Etzelmüller, 2009).

Os atributos do terreno usualmente são divididos em primários e secundários, também denominados compostos. Os primários são calculados diretamente do modelo digital do terreno e incluem medidas morfométricas como declividade, orientação da superfície, várias definições de curvatura, comprimento de fluxo (extensão do escoamento superficial) e acúmulo de fluxo (área de contribuição/captação a montante). 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 compostos, tais como o índice de umidade topográfica, o índice de poder de escoamento e o índice de posição topográfica, entre outros (Moore et al., 1991; Moore et al., 1993, Florinsky et al., 2002; De Reu et al., 2013).

Além da obtenção de medidas morfométricas, a modelagem digital do terreno também oferece possibilidades para automatizar o mapeamento de formas de relevo a fim de delimitar superfícies com características topográficas mais uniformes, tal como na interpretação de fotografias aéreas estereoscópicas. O interesse na classificação de formas reside no seu potencial para sintetizar a predominância de certos processos pedogenéticos e assim auxiliar a explicar a gênese e distribuição dos solos. Isso é útil tanto para o desenvolvimento de modelos preditivos solo-paisagem quanto para guiar novas amostragens em campo a fim de aperfeiçoar a representatividade e evitar coletas desnecessárias. As abordagens vão desde a segmentação de grandes compartimentos do relevo, tais como planícies, vales, encostas e terraços, até a individualização de áreas elementares que se aproximam do conceito de superfícies geomórficas (Hengl & Rossiter, 2003; MacMillan, 2003; Chiverrel et al, 2008; Li et al., 2011; Evans, 2012).

Uma das propostas mais bem sucedidas de classificação automatizada de formas do relevo foi implementada por MacMillan (2003) em um conjunto de programas chamado *LandMapr toolkit*©. A configuração padrão dos programas resulta em um mapa com 15 classes de formas, que expressam o padrão topográfico da superfície e sua posição relativa na paisagem. O algoritmo usa cinco atributos do terreno e emprega funções *fuzzy* para agrupar *pixels* segundo medidas morfométricas e posição relativa entre o topo e a base da encosta. O ponto de partida é a identificação dos talvegues e dos divisores de água, cujos critérios podem ser modificados pelo usuário para controlar o detalhe do mapeamento. O usuário também pode modificar as regras de classificação a fim de excluir classes inexistentes na sua área de estudo ou incluir novas classes, bem como alterar os limites das respectivas funções *fuzzy*. Dessa forma, os mesmos conceitos de classificação podem ser adaptados a diferentes padrões de relevo (Li et al., 2011).

2.3.2. A importância de dados legados de solos

Existe uma gama de fontes a partir das quais é possível obter covariáveis para descrever a maior parte dos fatores no modelo *s.c.o.r.p.a.n*, salvo o solo. Informações sobre classes ou propriedades dos solos em alguns locais são necessárias para ajustar o modelo de predição e avaliar a qualidade dos resultados, mas sua coleta *in loco* implica em dificuldades logísticas e financeiras equivalentes às dos levantamentos convencionais. Uma das alternativas para suprir as informações necessárias sem refazer levantamentos caros e extensivos é extraí-las de mapas e perfis de solo de levantamentos já existentes, o que tem sido uma estratégia amplamente adotada. Nesse contexto, dados existentes sobre os solos em certas áreas passaram a ser denominados de “dados legados”. Além de seu valor intrínseco, eles assumem relevância ainda maior em função das novas informações que podem ajudar a produzir (McBratney et al., 2003; Grunwald, 2006; Hengl et al., 2007; Sanchez, et al., 2009; Minasny & McBratney, 2010; Yang et al., 2011).

Minasny & McBratney (2010) propuseram um roteiro para definir estratégias de uso de dados legados na predição de propriedades, mas o esquema pode ser considerado válido também para classes de solo (Figura 1). O primeiro passo consiste na definição da área de interesse. Em seguida, tendo em vista o objetivo (classes ou propriedades), devem ser coletadas e

estruturadas covariáveis ambientais disponíveis e que possam descrever satisfatoriamente alguns dos fatores *s.c.o.r.p.a.n.* Após a estruturação das covariáveis, os dados existentes sobre solos devem ser reunidos e avaliados para então definir a estratégia de uso mais adequada em função do seu tipo, cobertura geográfica e qualidade. Obter medidas sobre a qualidade é fundamental porque mapas convencionais de solos contêm uma série de imperfeições e incertezas inerentes à subjetividade e às simplificações necessárias à sua elaboração, as quais frequentemente são ignoradas (Heuvelink & Webster, 2001; Carré et al., 2007; Malone et al, 2012; Miller, 2012; Phillips, 2013).

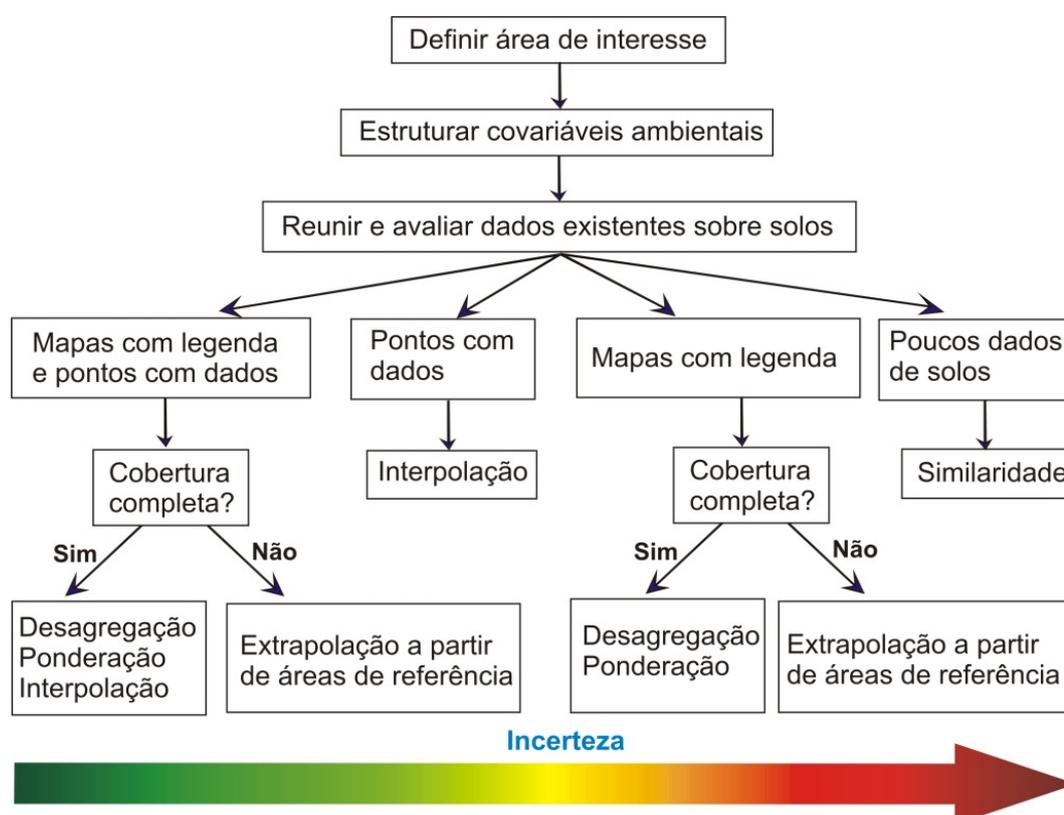


Figura 1. Estratégias para desenvolvimento de modelos preditivos a partir de dados legados de solos (Minasny & McBratney, 2010, adaptado).

Ainda antes da emergência do mapeamento digital, Beckett & Burrough (1971) e Bie & Beckett (1973) já propunham avaliar a qualidade de mapas de solos a partir da sua legenda, da exatidão na localização espacial das unidades de mapeamento, e da sua pureza e uniformidade interna. Forbes et al. (1987) e Rossiter (2000) introduziram medidas derivadas da escala e da textura do mapa, da sua legenda, da qualidade da base cartográfica, e de dados observados em campo. Hengl & Husnjak (2006) propuseram vários

indicadores para avaliar mapas legados no contexto de mapeamento digital, alguns dos quais computados a partir dos mapas digitalizados, outros a partir dos respectivos relatórios de levantamento e alguns a partir de dados de referência coletados em campo. Em geral, medidas derivadas dos mapas, das legendas e dos respectivos relatórios podem fornecer subsídios suficientes sobre sua qualidade a fim de prevenir o uso inadequado.

Uma vez avaliados, dependendo do tipo, cobertura geográfica e qualidade dos dados existentes, a estratégia de uso mais adequada pode então ser definida. A situação mais favorável é quando a área de interesse dispõe de mapas com legendas e pontos com dados analíticos e/ou descritivos (perfis e pontos de observação), e à medida que os dados forem menos completos a incerteza nos resultados da predição tende a aumentar. Se cobrirem toda a área de interesse, eles podem ser empregados para ponderação, desagregação ou interpolação espacial; se tiverem apenas cobertura parcial a extrapolação pode ser a opção. Em cada caso, medidas de qualidade podem auxiliar a definir a estratégia mais apropriada. Por exemplo, a desagregação pode ser viável se a escala efetiva do mapa for inferior à nominal e se as unidades de mapeamento forem predominantemente combinadas, com diferentes arranjos de classes taxonômicas. Por outro lado, para extrapolação convém que a escala cartográfica efetiva seja próxima da nominal e que predominem unidades de mapeamento simples (Hengl & Husnjak, 2006; Sarmiento et al., 2014).

A desagregação espacial busca descrever a variação espacial dos solos que não está explícita nas unidades de mapeamento de mapas convencionais. O objetivo é produzir novos mapas *raster* com a alocação das classes ou propriedades dos solos à sua respectiva posição na paisagem (Nauman et al., 2014; Subburayalu et al., 2014). No caso de classes de solo, sua individualização dentro das unidades de mapeamento pode ser feita tanto retendo as bordas (Bui & Moran, 2001; Thompson et al., 2010; Häring et al., 2012) quanto eliminando os limites dos polígonos originais (de Bruin et al., 1999; Bui & Moran, 2003; Hansen et al., 2009; Wei et al., 2010; Smith et al., 2012; Nauman & Thompson, 2014; Subburayalu et al., 2014; Odgers et al., 2014a). Em ambos os casos, os mapas desagregados tendem a ser consideravelmente mais detalhados que os mapas convencionais e

frequentemente revelam classes originalmente não mapeadas, ou seja, inclusões. Ao individualizar classes, a desagregação também torna possível produzir novos mapas descrevendo a variação espacial contínua de propriedades dos solos que compõem as unidades de mapeamento (Kerry et al., 2012; Nauman et al., 2012; Odgers et al., 2014b; Malone et al., 2014).

A extrapolação propõe o uso de dados de solos existentes em uma área representativa das relações solo-paisagem para criar regras de mapeamento com base em covariáveis ambientais disponíveis, e então prever classes ou propriedades de solos em áreas não mapeadas com características semelhantes (Fawcett et al., 1993; Lagacherie et al., 1995; Bui & Moran, 2001; Grinand et al., 2008; Lemercier et al., 2012; Cambule et al., 2013; Höfig et al., 2014). Esse processo adota uma lógica similar à usada em levantamentos convencionais, onde observações são feitas em alguns pontos representativos no terreno e depois são usadas como base para inferir e delinear a ocorrência dos solos na paisagem em toda a área de interesse. Uma vantagem da extrapolação com técnicas de mapeamento digital é a consistência, pois uma vez estabelecidas regras de mapeamento a partir de covariáveis disponíveis, elas são aplicadas uniformemente sobre toda a área a ser mapeada, em uma análise *pixel a pixel* (Zhu, 2000; Bui, 2004; Hempel et al., 2008; Minasny & McBratney, 2010).

2.3.3. Avaliação da qualidade de mapas digitais de solos

2.3.3.1. Validação com dados de campo

A forma clássica de avaliar a qualidade dos resultados em mapeamento digital de solos é comparar as predições com medidas de referência obtidas em um conjunto de pontos aleatoriamente distribuídos na área de estudo. No caso de mapas preditivos quantitativos, relativos a certas propriedades do solo, os valores preditos e de referência em cada ponto são comparados para calcular uma série de parâmetros que expressam o ajuste ou o desvio da predição, tais como o erro absoluto e proporcional, o erro médio quadrado (RMS), o coeficiente de correlação, intervalos de confiança e vários outros indicadores estatísticos (MacMillan, 2008).

Em mapas preditivos de classes, por outro lado, o conceito de ajuste ou de desvio não se aplica porque os valores preditos não representam uma grandeza. Eles são qualitativos e, neste caso, a comparação da classe predita

com a classe observada em cada ponto de referência admite apenas duas situações: concordante ou discordante. No entanto, indicadores quantitativos podem ser obtidos a partir da tabulação de contagens das duas situações em matrizes de erro, também denominadas de matrizes de confusão ou matrizes de contingência. Uma matriz de erro é uma tabela de dupla entrada na qual as classes de referência são plotadas ao longo das colunas e as classes preditas ao longo das linhas (Congalton, 1991; Foody, 2002; Pontius, 2000).

Em cada célula da matriz de erro são quantificadas as contagens para a respectiva combinação entre dados preditos e de referência, geralmente em percentagem do total de casos. Assim, as células da diagonal principal correspondem às concordâncias entre classes preditas e de referência, e sua soma fornece a percentagem total de acertos, também denominada exatidão geral ou acurácia geral. As concordâncias ao longo de cada linha em relação ao total da linha expressam a acurácia do usuário, ou seja, a proporção da classe predita que concorda com os dados de referência, enquanto as discordâncias em relação ao mesmo total expressam os erros de inclusão, isto é, a proporção de outras classes preditas como a classe em questão. As concordâncias ao longo de cada coluna em relação ao total da coluna expressam a acurácia do mapeador, ou seja, a proporção da classe de referência que foi corretamente predita, enquanto as discordâncias em relação ao mesmo total expressam os erros de omissão, isto é, a proporção da classe de referência que foi atribuída a outras classes (Campbell & Mortenson, 1989; Rossiter, 2001; Foody, 2002; Filippi & Jensen 2006).

A partir da matriz de erro também pode ser calculado o índice *kappa* (Cohen, 1960), uma estatística que procura excluir a porção das concordâncias que pode ser atribuída ao acaso. No entanto, o uso do índice *kappa* para caracterizar a acurácia de mapas de classes tem sido muito questionado, especialmente pela tendência em subestimar a qualidade do mapa avaliado ao remover concordâncias que supostamente são devidas ao acaso. Além disso, ele não leva em conta discordâncias que podem ser resultantes de problemas posicionais, situação muito comum na comparação de dados georreferenciados de diferentes origens. Em muitos casos a imprecisão nas coordenadas dos pontos de referência resulta na sua alocação a *pixels* incorretos do mapa predito. O próprio particionamento discreto do espaço em *pixels* causa

discordâncias porque um ponto de referência pode cair em certo *pixel* simplesmente em função da resolução espacial utilizada (Pontius, 2000; Rossiter, 2001; Foody, 2002; Grinand et al., 2008; MacMillan, 2008).

Pontius & Millones (2011) também mostraram que existem vários cálculos possíveis para o índice *kappa*, e resultados contraditórios. Os autores sugeriram seu abandono e propuseram duas novas medidas: a discordância de quantidade (DQ) e a discordância de alocação (DA). A primeira expressa a fração da diferença entre os dados preditos e os de referência devida ao ajuste imperfeito nas proporções das classes, enquanto a segunda mede a fração da diferença devida ao ajuste imperfeito na alocação espacial das classes, dadas suas proporções nos dados preditos e nos de referência. Enquanto *kappa* tenta medir quanto a concordância é melhor que a obtida ao acaso, DQ e DA medem quanto ela é menos que perfeita. Elas decompõem a discordância em dois componentes relacionados à proporção e à alocação espacial das classes preditas, fornecendo informação adicional que ajuda a explicar o erro. Para uma mesma discordância total, DA maior que DQ é preferível porque significa que parte de algumas classes foram preditas nas vizinhanças dos pontos de referência, mas que sua superfície se aproxima da real.

A compensação ou captura de incertezas posicionais também pode ser feita pelo uso de raios de busca (*buffers*) em torno dos pontos de referência, o que possibilita ainda detectar a transição gradual entre classes que ocorre na realidade. É também uma alternativa para aumentar o número de pontos de referência com base na autocorrelação espacial, presumindo-se que locais muito próximos tendem a pertencer à mesma classe de solo ou a ter propriedades similares. O conhecimento local sobre as relações solo-paisagem pode ser usado para limitar o raio de busca a distâncias apropriadas em cada caso (Smith et al., 2012; Nauman & Thompson, 2014).

2.3.3.2. Outras formas de validação

A validação com pontos de referência coletados em campo enfrenta dificuldades, pois raramente existem pontos disponíveis em quantidade e com distribuição espacial adequada, e sua coleta *in loco* frequentemente é inviável em função de custos e de outras limitações. Dessa forma, no contexto de mapeamento digital qualquer medida potencialmente útil para caracterizar a qualidade de mapas preditos passou a ter relevância. Até mesmo a inspeção

visual, menosprezada por muito tempo em detrimento de medidas mais objetivas, passou a ser reconsiderada frente à escassez de dados de referência, valorizando assim o conhecimento local sobre os solos (Hansen et al., 2009; Li et al., 2012).

O conhecimento tácito sobre os solos de determinada área também pode servir de base para avaliações quantitativas da qualidade da predição mediante a execução de testes lógicos contra os valores de algumas covariáveis ambientais usando abordagens conceituais como plausibilidade, por exemplo (Möller et al., 2012). Informações descritivas ou quantitativas extraídas de mapas convencionais de solos, pontos de campo ou de relatórios de levantamentos de solos, tais como número de classes por unidade de mapeamento, fração ou área ocupada por classe, também são indicativos válidos para avaliar a qualidade das predições (Bui & Moran, 2001; Roecker et al, 2010; Odgers et al., 2014a).

A evolução dos recursos de processamento tem despertado ainda crescente interesse em medidas sobre a incerteza nas predições. Uma das formas de expressá-la é pela preparação e apresentação de mapas dos resíduos resultantes do processo preditivo. Outra possibilidade é produzir múltiplas realizações de cada mapa preditivo mediante a variação da amostragem, das variáveis de entrada ou do próprio método de predição. Em mapas quantitativos a dispersão dos resultados obtidos com as múltiplas realizações pode fornecer um indicativo da incerteza ou da confiança nas predições em cada local. No caso de mapas preditivos de classes, a frequência relativa da classe predita em cada *pixel* ao longo do conjunto de mapas de preditivos obtidos com múltiplas realizações pode ser uma aproximação para medir a confiança (Bui & Moran, 2001; Nauman & Thompson, 2014).

3. CAPÍTULO II - CARACTERIZAÇÃO DE MAPAS LEGADOS DE SOLOS: USO DE INDICADORES EM MAPAS COM DIFERENTES ESCALAS NO RIO GRANDE DO SUL

3.1. Introdução

Um levantamento de solos é uma estratégia científica com base nos conceitos de fatores de formação de solos associados com relações solo-paisagem (Hudson, 1992). Os levantamentos convencionais contemplam o estudo do terreno e das características de perfis representativos, compreendendo descrição morfológica, caracterização física e química, classificação taxonômica e espacialização (mapeamento) da ocorrência dos diferentes tipos de solos. Assim, os mapas de solos são a materialização de um conhecimento estruturado sobre a distribuição dos solos na paisagem (Bui, 2004).

Os levantamentos de solos são enquadrados em diferentes níveis de detalhamento, cada qual com recomendações para a densidade de observações e amostragens em campo e para a escala de apresentação do mapa final. Os dois primeiros admitem variações em função das características da área e da experiência da equipe e podem ser substancialmente reduzidos com o uso de geotecnologias (IBGE, 2007). A escala, contudo, costuma ser determinada em função dos objetivos do levantamento e do material usado como base para delimitar as unidades de mapeamento, dentre outros fatores. Todavia, sendo um processo manual, a delimitação dificilmente segue normas cartográficas estritas quanto à escala escolhida e aos aspectos a ela relacionados, como a acurácia posicional, por exemplo (Hengl & Husnjak, 2006; Nussbaum et al., 2011).

De fato, o mapeamento enfrenta o desafio de transcrever a variação tridimensional contínua dos solos e de suas propriedades na paisagem para fronteiras bidimensionais discretas que delimitam a ocorrência de classes taxonômicas. Há também a dificuldade prática de manter uniformidade durante

o processo, evitando a criação de polígonos com dimensão inferior à área mínima mapeável em algumas partes do mapa e generalização excessiva em outras. Como consequência, mapas convencionais de solos inevitavelmente contêm uma série de imperfeições e incertezas, decorrentes tanto de limitações inerentes à própria forma de representação cartográfica quanto de eventuais falhas ou lacunas nas amostragens em campo (Heuvelink & Webster, 2001; Malone et al, 2012; Miller, 2012; Phillips, 2013).

Quanto mais detalhado for o levantamento de solos, menores tendem a ser as imperfeições e incertezas inerentes ao mapa e mais homogêneas tendem a ser as unidades de mapeamento. Em contrapartida, o esforço, o tempo e o custo para a execução das atividades de campo, as análises em laboratório e a elaboração do próprio mapa crescem consideravelmente (Nussbaum et al., 2011). Devido a esses aspectos, em contraste com as áreas a serem cobertas, a maioria dos levantamentos existentes foi executada com baixo nível de detalhamento, resultando em mapas de solos em escalas pequenas (McBratney et al., 2003; Sanchez et al., 2009). Todavia, apesar das limitações, eles ainda constituem o melhor, senão o único, registro disponível sobre a ocorrência e distribuição dos solos em muitos locais, e mais recentemente passaram a receber a denominação de “mapas legados” (Carré et al., 2007).

Nos últimos anos, dados legados de solos têm adquirido importância crescente como entrada para diversas aplicações, com destaque para estudos em mapeamento digital de solos e em modelagem ambiental (Omuto et al., 2013). Frequentemente os polígonos dos mapas convencionais de solos são convertidos do modelo vetorial para o formato de grade (*raster*) em SIG, sendo a resolução espacial definida em função de outros dados *raster* disponíveis, como o SRTM (*Shuttle Radar Topography Mission*) ou imagens NDVI/MODIS (*Normalized Difference Vegetation Index/Moderate Resolution Imaging Spectroradiometer*). No entanto, essa resolução pode resultar excessivamente fina ou grosseira para o detalhe espacial e temático que o mapa de solos efetivamente possui (Nussbaum et al., 2011; Miller, 2012).

Em mapeamento digital de solos, a resolução espacial pode ser usada como recurso tanto para incrementar o detalhe quanto para generalizar mapas convencionais existentes, com métodos que possibilitem controlar o

processo e obter resultados consistentes (Malone et al., 2012; Nauman & Thompson, 2014). Em outras aplicações, entretanto, especificações inadequadas para a resolução espacial e outros parâmetros podem violar pressupostos e gerar resultados enganosos, levando a interpretações equivocadas. Assim, conhecer as características do mapa convencional torna-se fundamental para orientar o uso correto de qualquer informação dele extraída ou derivada (Carré et al., 2007; Basayigit & Senol, 2008).

A preocupação quanto à avaliação de mapas de solos para suprir informações a diferentes usos existe há várias décadas. Beckett & Burrough (1971) propuseram avaliar a qualidade dos mapas a partir da legenda, da exatidão da localização espacial dos limites dos polígonos e da pureza e uniformidade das unidades de mapeamento, a última modificada posteriormente por Bie & Beckett (1973). Forbes et al. (1987) e Rossiter (2000) introduziram o conceito de adequação, considerando: escala e textura do mapa; legenda; qualidade do mapa base; e verdade de campo, ou acurácia temática. Hunter et al. (2003) discutiram a “usabilidade” de dados geoespaciais em um contexto mais amplo, considerando basicamente: qualidade dos dados (acurácia, completeza, consistência lógica); forma(to); acessibilidade/preço, e qualidade dos metadados.

Hengl & Husnjak (2006) apresentaram a diferença entre os conceitos de qualidade, adequação e usabilidade. A primeira envolve um conjunto de características eminentemente técnicas, a segunda depende do contexto do problema a ser resolvido, e a terceira considera os dois em relação à satisfação das necessidades do usuário. Com base nessas abordagens, os autores propuseram 10 indicadores, com graus variados de complexidade, computados a partir dos mapas digitalizados, dos respectivos relatórios de levantamento e de dados de referência coletados em campo.

O objetivo deste estudo foi descrever e comparar um conjunto de mapas convencionais de solos do Rio Grande do Sul usando indicadores quantitativos computados diretamente a partir dos mapas digitalizados e de suas legendas. A hipótese é que a escala de apresentação ou o nível de detalhe do levantamento não são suficientes para caracterizar a real qualidade de mapas legados de solos. Pressupõe-se que tais indicadores forneçam medidas objetivas sobre as características geométricas e temáticas dos mapas

existentes, que sirvam de subsídio para a tomada de decisão consciente sobre seu uso.

3.2. Material e métodos

A área de estudo compreende a totalidade do Estado do Rio Grande do Sul e algumas áreas específicas de seu território, onde existem informações de levantamentos de solos executados no passado. As coordenadas extremas situam-se entre as latitudes 27°05'S e 33°45'S e as longitudes 49°41'O e 57°38'O (Figura 2). Predomina no Estado o clima subtropical úmido com verões quentes (Cfa, conforme Köppen) e com verões amenos (Cfb) nas porções mais elevadas. A geologia compreende uma diversidade de litologias que incluem rochas graníticas do Eon Arqueano, areníticas e basálticas da Era Mesozóica e depósitos sedimentares da Era Cenozóica. O relevo varia de plano a escarpado e apresenta altitudes que vão de menos de 10 m a mais de 1.000 m. Essas litologias, altitudes e declividades integradas geram diversas formas do relevo, as quais correspondem a várias unidades geomorfológicas (IBGE, 1986). A vegetação é representada por florestas ombrófilas e estacionais e por formações campestres arbustivas e graminosas.

O material utilizado neste estudo compreende nove mapas de solos disponíveis em meio digital em formatos compatíveis com uso em SIG. Os mapas foram produzidos por meio de levantamentos convencionais em diferentes épocas e níveis de detalhamento, cobrem uma variedade de extensões geográficas e paisagens e foram elaborados em quatro escalas de apresentação (Tabela 1 e Figura 2). O processamento e análise dos mapas digitalizados e georreferenciados foi realizado com auxílio do *software* ArcGIS (ESRI, 2013).

Inicialmente, calcularam-se a área e o perímetro de cada polígono e, a partir dessas medidas, determinaram-se a escala cartográfica efetiva dos mapas de solos e o índice de complexidade de forma. Segundo Hengl & Husnjak (2006), a escala efetiva é definida como uma medida objetiva que indica se o entrelaçamento dos polígonos no mapa de solos corresponde à sua escala nominal, a escala de apresentação. A escala efetiva foi calculada pela seguinte equação (Forbes et al., 1987):

$$ESN = NSN \times (IMR/2)$$

em que ESN é o divisor da escala efetiva; NSN, o divisor da escala nominal; e IMR, o índice de máxima redução, o fator pelo qual a escala do mapa pode ser reduzida até que o tamanho médio dos polígonos delimitados (ASD) seja igual à área mínima mapeável (MLD), que no Brasil é especificada como 0,4 cm² (IBGE, 2007). O valor de MLD foi previamente convertido de cm² para seu tamanho equivalente em ha, considerando a escala nominal de cada mapa.

ASD e *IMR* foram calculadas pelas seguintes equações (Forbes et al., 1987):

$$ASD = \sum_{j=1}^m A_j / m$$

$$IMR = \sqrt{ASD/MLD}$$

em que A_j é a área do polígono j (em ha), e m o número total de polígonos no mapa.

Tabela 1. Identificação, local, área mapeada, escala nominal e fonte de mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.

Mapa	Local/abrangência	Área mapeada		Escala nominal	Fonte
		ha	km ²		
1	Vale dos Vinhedos	8.121,6	81,2	1:10.000	Flores et al. (2012)
2	Município de Porto Alegre	47.561,1	475,6	1:50.000	Schneider et al. (2008)
3	Folha Palomas ^{/1}	62.621,1	626,2	1:50.000	Flores et al. (2007a)
4	Folha Pinheiro Machado ^{/1}	65.779,1	657,8	1:50.000	Flores et al. (2007b)
5	Folha Encruzilhada do Sul ^{/1}	66.418,9	664,2	1:50.000	Flores et al. (2009)
6	Bacia do rio Gravataí	201.253,0	2.012,5	1:50.000	Jungblut (1994)
7	Serra Gaúcha	1.348.961,3	13.489,6	1:50.000	Flores et al. (2007c)
8	Rio Grande do Sul	28.204.501,6	282.045,0	1:250.000	SAA/IBGE (2003)
9	Rio Grande do Sul	28.204.501,6	282.045,0	1:750.000	Machado & Fitz (2001)

^{/1} Nome da folha de acordo com a articulação das cartas topográficas da Diretoria de Serviço Geográfico (DSG) do Exército na escala 1:50.000.

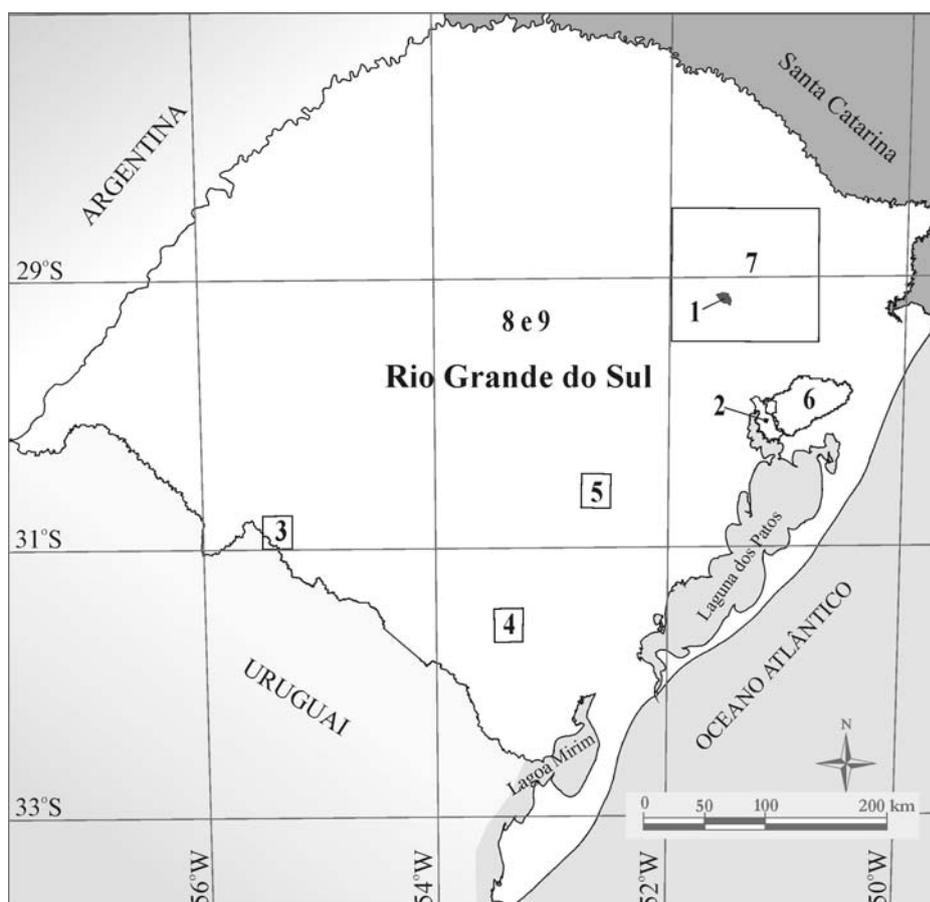


Figura 2. Localização de mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000 (1 = Vale dos Vinhedos; 2 = Município de Porto Alegre; 3 = Folha Palomas; 4 = Folha Pinheiro Machado; 5 = Folha Encruzilhada do Sul; 6 = bacia do rio Gravataí; 7 = Serra Gaúcha; 8 e 9 = Rio Grande do Sul).

O índice de complexidade de forma (S) procura descrever a geometria geral dos polígonos nos mapas de solos, e consiste em uma relação entre perímetro e área de cada polígono, que foi calculada pela seguinte equação (Hole, 1978):

$$S = P/2\pi r, \text{ onde } r = \sqrt{A/\pi}$$

em que P o perímetro do polígono; A , a área do polígono; e r , o raio de um círculo com área igual à área do polígono em questão, todas expressas em unidades de grandeza equivalente; neste caso, m^2 para A e m para P e r . Valores de S próximos a 1 caracterizam polígonos bastante compactos e simples, enquanto os mais distantes correspondem a formas mais complexas (polígonos mais estreitos ou ramificados), o que em geral significa acurácia posicional e escala efetiva maiores (D'Avello & McLeese, 1998).

Por último, a partir da legenda de cada mapa, quantificaram-se o número de classes taxonômicas (do primeiro ao quarto nível) e o número de tipos de solos, convencionando-se “tipo de solo” como a descrição completa, incluindo as características ou propriedades que diferenciam os solos após o quarto nível de classificação taxonômica, tais como fases de relevo e pedregosidade, atividade de argila e tipo de horizonte superficial, entre outros. Adicionalmente, calcularam-se também o número total de unidades de mapeamento, o número de unidades simples e combinadas e a fração da área total representada por dados não uniformes, pela seguinte equação (Nussbaum et al., 2011):

$$P_{ND} = (A_{ND}/A_M) \times 100$$

em que P_{ND} é a fração com dados não uniformes; A_{ND} , a área ocupada por dados não uniformes (unidades de mapeamento combinadas) e A_M , a área total mapeada.

3.3. Resultados e discussão

A escala efetiva calculada em todos os mapas foi inferior à sua escala nominal (Tabela 2). Curiosamente, as menores diferenças entre as duas escalas ocorreram nos mapas 1 e 9, cujas escalas de apresentação são, respectivamente, a maior e a menor do conjunto de mapas utilizado (Inicialmente, calcularam-se a área e o perímetro de cada polígono e, a partir dessas medidas, determinaram-se a escala cartográfica efetiva dos mapas de solos e o índice de complexidade de forma. Segundo Hengl & Husnjak (2006), a escala efetiva é definida como uma medida objetiva que indica se o entrelaçamento dos polígonos no mapa de solos corresponde à sua escala nominal, a escala de apresentação. A escala efetiva foi calculada pela seguinte equação (Forbes et al., 1987):

$$ESN = NSN \times (IMR/2)$$

em que ESN é o divisor da escala efetiva; NSN, o divisor da escala nominal; e IMR, o índice de máxima redução, o fator pelo qual a escala do mapa pode ser reduzida até que o tamanho médio dos polígonos delimitados (ASD) seja igual à área mínima mapeável (MLD), que no Brasil é especificada como 0,4 cm² (IBGE, 2007). O valor de MLD foi previamente convertido de cm² para seu tamanho equivalente em ha, considerando a escala nominal de cada mapa.

ASD e *IMR* foram calculadas pelas seguintes equações (Forbes et al., 1987):

$$ASD = \sum_{j=1}^m A_j / m$$

$$IMR = \sqrt{ASD/MLD}$$

em que A_j é a área do polígono j (em ha), e m o número total de polígonos no mapa.

Tabela 1). Ambos também apresentaram o menor *IMR*, com valores igualmente próximos, o que demonstra que, embora elaborados em escalas de apresentação muito distantes, os dois mapas possuem consistência semelhante quanto à generalização, com poucos polígonos excessivamente grandes. Destaca-se também o mapa 8, cuja escala nominal é numericamente três vezes maior que a do mapa 9; entretanto, sua escala efetiva resultou muito próxima da obtida para o último. O *IMR* alto confirma que o primeiro é consideravelmente mais generalizado do que faz supor sua escala nominal.

Entre os mapas elaborados na escala 1:50.000, o mapa 2 apresentou a escala efetiva mais próxima da nominal enquanto os mapas 6 e 7 apresentaram as maiores discrepâncias entre as duas escalas. O *IMR* evidencia que o tamanho dos polígonos nestes dois mapas está acima do ideal para sua escala de apresentação, o que pode ser produto tanto de simplificação das unidades de mapeamento quanto de dificuldade de individualização de solos com ocorrência muito intrincada. Os três mapas restantes apresentaram diferenças com magnitude semelhante entre a escala efetiva e a nominal, com valores de *IMR* igualmente próximos entre si.

Todos os mapas apresentaram polígonos com dimensão inferior à área mínima mapeável (*MLD*) para sua escala nominal, com o maior número no mapa 1, seguido dos mapas 8 e 9. Entre os mapas na escala nominal 1:50.000, o mapa 2 apresentou a maior frequência de casos, em parte devido à delimitação de Tipos de Terreno, que são áreas geralmente pequenas. O mapa 4 discriminou a menor quantidade de polígonos menores que a área mínima mapeável, e os demais oscilaram em torno de valores intermediários próximos.

O menor e o maior índice de complexidade de forma (*S*) foram obtidos respectivamente para os mapas 1 e 8 (Tabela 2). A *ASD* pequena e os polígonos inferiores à *MLD* explicam o baixo *S* do primeiro, pois ambos tendem

a formas mais simples. De outro lado, o desvio padrão de S evidencia certa homogeneidade na delimitação dos polígonos em todo o mapa, provavelmente em razão do uso de geotecnologias (Sarmiento et al., 2008). No segundo, uma inspeção visual evidenciou que os polígonos possuem bordas bastante sinuosas e recortadas, o que resultou no maior S, apesar da ASD grande. O contrário foi observado no mapa 9, cujos polígonos têm bordas bem mais simplificadas, mesmo sendo sua escala efetiva próxima à do anterior. Entre os mapas 1:50.000, o mapa 2 apresentou o menor S, denotando maior simplificação que os demais, que apresentaram S próximo entre si. Todavia, segundo a classificação proposta por Hole (1978), todos os mapas se enquadraram na mesma categoria: polígonos com forma moderadamente complexa.

Considerando-se o conjunto de mapas utilizados (Tabela 3), o número de classes taxonômicas e de tipos de solos foi aproximadamente proporcional à escala e à extensão da área mapeada. No caso dos mapas 8 e 9, o número de classes taxonômicas e de tipos de solos foram praticamente idênticos, o que é coerente com suas escalas efetivas próximas (Tabela 2). O mapa 6 se diferenciou dos demais por apresentar classes taxonômicas apenas até o terceiro nível, o que pode restringir certos usos por causa da impossibilidade de inferir determinadas propriedades dos solos.

Tabela 2. Escala nominal, número de polígonos, área média dos polígonos (ASD), área mínima mapeável (MLD), índice de máxima redução (IMR), escala efetiva, número de polígonos menores que a área mínima mapeável (Políg.<MLD) e parâmetros estatísticos do índice de complexidade de forma (S) para mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.

Mapa	Abrangência	Escala nominal	Núm. de políg.	ASD (ha)	MLD (ha)	IMR	Escala efetiva	Políg. <MLD	Índice de complexidade de forma (S)			
									Média	Mínimo	Máximo	Desvio padrão
1	Vale dos Vinhedos	1:10.000	1648	4,9	0,4	3,51	1:17.550	223	1,87	1,02	13,72	0,96
2	Município de Porto Alegre	1:50.000	228	208,8	10	4,57	1:114.229	41	2,07	1,06	11,16	1,25
3	Folha Palomas ¹	1:50.000	159	393,8	10	6,28	1:156.892	22	2,51	1,01	11,92	1,94
4	Folha Pinheiro Machado ¹	1:50.000	138	476,7	10	6,90	1:172.601	3	2,54	1,03	6,13	1,14
5	Folha Encruzilhada do Sul ¹	1:50.000	195	340,6	10	5,84	1:145.904	16	2,46	1,02	7,93	1,24
6	Bacia do rio Gravataí	1:50.000	126	1.597,3	10	12,64	1:315.955	15	2,40	1,04	12,67	2,01
7	Serra Gaúcha	1:50.000	1626	829,6	10	9,11	1:227.709	24	2,48	1,01	22,58	1,61
8	Rio Grande do Sul	1:250.000	1214	21.812,1	250	9,34	1:1.167.586	196	2,69	1,04	20,26	1,78
9	Rio Grande do Sul	1:750.000	832	32.126,5	2.250	3,78	1:1.417.006	94	2,03	1,04	12,17	1,04

¹ Nome da folha de acordo com a articulação das cartas topográficas da Diretoria de Serviço Geográfico (DSG) do Exército na escala 1:50.000.

Tabela 3. Escala nominal, classes taxonômicas em diferentes níveis de classificação (N1 = Ordem, N2 = Subordem, N3 = Grande Grupo, N4 = Subgrupo), tipos de solos, unidades de mapeamento e fração da área mapeada com dados não uniformes (PND) em mapas convencionais de solos do Rio Grande do Sul, elaborados nas escalas 1:10.000 a 1:750.000.

Mapa	Abrangência	Escala nominal	Classes taxonômicas				Tipos de solos	Unidades de mapeamento		PND (%)	
			N1	N2	N3	N4		Total	Simples Combinadas		
1	Vale dos Vinhedos	1:10.000	6	12	37	54	156	156	0	0,00	
2	Município de Porto Alegre	1:50.000	6	10	13	15	15	12	2	10	98,31
3	Folha Palomas ¹	1:50.000	5	8	11	12	20	16	12	4	29,16
4	Folha Pinheiro Machado ¹	1:50.000	3	6	8	8	41	34	14	20	66,70
5	Folha Encruzilhada do Sul ¹	1:50.000	3	6	8	8	45	31	14	17	45,55
6	Bacia do rio Gravataí	1:50.000	7	9	12	---	27	12	5	7	83,06
7	Serra Gaúcha	1:50.000	7	11	21	24	92	63	9	54	97,32
8	Rio Grande do Sul	1:250.000	13	23	37	44	45	183	46	137	80,14
9	Rio Grande do Sul	1:750.000	13	25	36	48	48	75	46	29	27,54

¹ Nome da folha de acordo com a articulação das cartas topográficas da Diretoria de Serviço Geográfico (DSG) do Exército na escala 1:50.000.

O mapa 8 apresentou o maior número de unidades de mapeamento, contrastando com o mapa 9, com menos de metade daquele. A diferença de unidades combinadas e PND entre ambos é ainda mais acentuada, o que revela estratégia distinta na delimitação da ocorrência dos solos, apesar de cobrirem a mesma área, terem escala efetiva semelhante e praticamente o mesmo número de classes taxonômicas e de tipos de solos. O mapa 1 foi o segundo mais numeroso em unidades de mapeamento por causa de todas serem unidades simples, representando uma única classe taxonômica. Excetuando-se esse e o mapa 3, nos demais predominaram unidades de mapeamento combinadas e a PND foi aproximadamente proporcional à quantidade daquelas. A PND mais alta foi encontrada no mapa 7, explicada pela ocorrência intrincada dos solos em razão da complexidade do relevo da região (Sarmiento et al., 2008).

Os mapas utilizados neste estudo constituem uma amostra reduzida dos produtos cartográficos de levantamentos convencionais de solos executados nas diversas regiões do Brasil. Eles foram utilizados pela simples razão de estarem acessíveis aos autores em formato digital compatível com o uso imediato em SIG, sem qualquer outro critério de seleção. No entanto, o conjunto é representativo da situação da massa de dados legados de solos no Brasil e em vários outros países: um mosaico de mapas em escalas variadas, cobrindo diferentes paisagens e extensões geográficas, elaborados em épocas diferentes e a partir de materiais básicos, metodologias e recursos tecnológicos distintos, bem como executados por equipes diversas.

Os resultados evidenciaram que mapas legados de solos tendem a apresentar qualidade inferior à presumida para sua escala de apresentação. Mesmo em mapas com a mesma escala nominal, a escala efetiva apresentou uma variação considerável. A magnitude das diferenças observadas foi similar às encontradas por Hengl & Husnjak (2006), em cartas de solos 1:50.000 de várias paisagens na Croácia, e por Geng et al. (2010), em um conjunto maior e mais diversificado de mapas de solos de províncias do Canadá. O enquadramento de todos os mapas na categoria de forma moderadamente complexa, independente da escala nominal, reproduz o que foi verificado por

Hole (1978), no Estado norte-americano de Wisconsin, e por Hengl & Husnjak (2006), na Croácia.

A proporcionalidade observada do número de classes taxonômicas e de tipos de solos com a escala e o tamanho da área mapeada é esperada, pois o incremento na escala de mapeamento exige maior intensidade de amostragem em campo, aumentando a chance de identificação de novas classes de solos. O mesmo se verifica com a expansão da área mapeada, por causa da maior diversidade de ambientes e de paisagens encontradas (Beckett & Burrough, 1971; Rossiter, 2000). Já o número de unidades de mapeamento, unidades combinadas e a PND são mais dependentes das especificidades de cada levantamento. O uso de unidades combinadas, por exemplo, tende a crescer com a redução da escala ou com o aumento da complexidade da paisagem, porque o delineamento de classes taxonômicas individualizadas torna-se mais difícil ou demorado (Dalmolin et al., 2004; Rossiter, 2004; Geng et al.; 2010). No entanto, apesar da mesma área de cobertura e da escala efetiva e número de classes taxonômicas similares, os mapas 8 e 9 apresentaram valores contraditórios para esses itens. A predominância de unidades simples no segundo faz presumir maior refinamento na delimitação da ocorrência dos solos, porém sua escala reduzida torna a individualização de classes taxonômicas inviável. O que ocorreu é que apenas a classe dominante foi identificada, as inevitáveis inclusões de outras classes não foram explicitadas.

Os resultados demonstram a importância e a necessidade de uma avaliação objetiva para orientar a tomada de decisões sobre o uso de mapas legados de solos. Mais do que desqualificar um determinado mapa, o cálculo de indicadores possibilita avaliar sua compatibilidade com os objetivos específicos de diferentes usuários (Rossiter, 2004, Malone et al., 2012). Conforme observado por Carré et al. (2007), isso é particularmente relevante em MDS, pois em geral as abordagens buscam otimizar a acurácia dos modelos e minimizar os erros de predição, sem a devida atenção à qualidade dos dados legados usados como entrada. A falta de avaliação prévia ou a *posteriori* de tais dados também é frequente em outras áreas de aplicação.

Dos indicadores utilizados, os relativos ao tamanho e forma dos polígonos têm a vantagem da facilidade e rapidez de determinação, salvo se os

mapas existirem apenas na forma impressa e necessitarem ser previamente estruturados em SIG (Dalmolin et al., 2004). A escala efetiva propicia leitura e interpretação direta, além de constituir uma informação especialmente útil para MDS porque tanto a captura da variabilidade espacial dos solos quanto a própria capacidade de predição de classes ou propriedades são dependentes da escala (Lin et al., 2005; Kim & Zheng, 2011). O índice de complexidade de forma também fornece informação de interesse para MDS porque polígonos de bordas mais detalhadas tendem a reduzir a inclusão de solos de unidades de mapeamento adjacentes e, conseqüentemente, a incerteza. Polígonos menores que a área mínima mapeável também tendem a ter menos inclusões de outros solos, apesar da incerteza proporcionalmente alta no traçado de seus limites (Rossiter, 2000).

Os indicadores temáticos obtidos da legenda são igualmente relevantes, mas necessitam de interpretação mais cuidadosa. O número de unidades de mapeamento combinadas e a PND, por exemplo, dependem da estratégia definida pelos mapeadores em função da complexidade da paisagem, do tamanho da área e do material cartográfico básico, em balanço com os recursos financeiros, o tamanho da equipe e o tempo. Conseqüentemente, cada levantamento acaba tendo características próprias (Beckett & Burrough, 1971; Bie & Beckett, 1973; Hengl & Husnjak, 2006; Basayigit & Senol, 2008; Geng et al., 2010). Assim, além de avaliar se um mapa é adequado para um determinado propósito, esses indicadores podem ser muito úteis para comparar mapas da mesma área ou de áreas adjacentes com vistas à sua sistematização e compatibilização (Odgers et al. 2014a).

Nesse sentido, tais indicadores evidenciaram que os mapas 2 e 6, que são adjacentes, requerem uma harmonização prévia de escala e de legenda para sua junção, haja vista as diferenças na escala efetiva, no nível de classificação taxonômica e no uso de unidades de mapeamento combinadas. Eles revelaram também que existe potencial para incrementar o detalhe espacial no mapa 8, pois o número de unidades de mapeamento, de unidades combinadas e PND mostram que sua riqueza temática é desproporcional à escala efetiva, em comparação com o mapa 9 que cobre a mesma área. O mesmo foi constatado anteriormente para o mapa 7 por Sarmiento et al. (2013). A desagregação de unidades de mapeamento combinadas empregando

modelos de MDS é uma alternativa a ser considerada para esses casos (Nauman & Thompson, 2014; Odgers et al., 2014a).

As medidas propostas neste estudo possibilitaram uma caracterização rápida e objetiva e, embora contemplem poucas características, têm potencial de aplicação prática na sistematização e documentação dos dados heterogêneos de solos disponíveis no Brasil. Isso se torna importante à medida que estruturas para catalogação e disponibilização de bases de dados geoespaciais via Internet, como a Infraestrutura Nacional de Dados Espaciais (INDE), gradativamente se consolidam (Rossiter, 2004). No entanto, apesar da vantagem de depender somente dos mapas e das respectivas legendas, e da interpretação relativamente simples, elas apenas descrevem certos aspectos de um mapa de solos. A avaliação propriamente dita depende do que o usuário pretende fazer com o mapa e, para alguns propósitos, tais medidas podem ser insuficientes.

Aquelas relativas ao tamanho e à forma dos polígonos, por exemplo, não expressam a incerteza no traçado dos limites nem eventuais deficiências no georreferenciamento, comuns em mapas anteriores à disseminação de programas de SIG, receptores GPS e imagens de satélite (Rukhovich et al., 2013). De outro lado, os indicadores derivados da legenda não expressam o quão diferentes realmente são as classes de solos, nem se os limites entre unidades de mapeamento fazem separações úteis, o que assume relevância à medida que classes distintas podem ter algumas propriedades bastante similares; por exemplo, se as classes de uma unidade de mapeamento combinada tiverem a mesma aptidão agrícola, o fato de elas não estarem individualizadas é irrelevante para tal avaliação. O mesmo é válido para classes em unidades de mapeamento diferentes. Nessas situações torna-se necessário usar medidas que possibilitem caracterizar o grau de similaridade entre classes (Minasny & McBratney, 2007; Phillips, 2013).

Existem inúmeros outros indicadores para mensurar uma diversidade de características espaciais e temáticas, além da similaridade entre classes. No entanto, muitos deles dependem da coleta de dados em campo, o que implica em disponibilidade de pessoal e de recursos, ou de métodos avançados de análise estatística, que não são de domínio da maioria dos potenciais usuários dos mapas (Rossiter, 2000; Hengl & Husnjak, 2006;

Nussbaum et al., 2011). No caso do Brasil, convém efetuar mais estudos a fim de recomendar indicadores de determinação viável e aplicação útil aos dados legados de solos.

3.4. Conclusões

- i. A qualidade de mapas convencionais de solos no Rio Grande do Sul tende a ser inferior à presumida, o que também é esperado em outras regiões do Brasil em virtude das normas e recomendações em comum;
- ii. As incertezas inerentes à delimitação dos polígonos e à forma de agregação das informações temáticas em unidades de mapeamento não devem ser negligenciadas;
- iii. A escala efetiva é uma medida simples, de interpretação direta e com grande impacto sobre as pretensões dos usuários e os resultados de modelos, sendo recomendável sua adoção rotineira para avaliar mapas convencionais de solos;
- iv. Indicadores calculados a partir dos polígonos e da legenda fornecem informações úteis, mas não completas; por isso é recomendável testar outras medidas, assim como a aplicação a mapas convencionais de solos de outras áreas.

4. CAPÍTULO III – APERFEIÇOAMENTO DE MAPAS LEGADOS DE SOLOS POR DESAGREGAÇÃO DE POLÍGONOS USANDO CONHECIMENTO ESPECIALISTA E TÉCNICAS DE MAPEAMENTO DIGITAL

4.1. Introdução

Os levantamentos de solos surgiram tendo como principal propósito atender questões voltadas ao planejamento da produção agrícola e do uso da terra. Ao longo do tempo, normas e recomendações foram criadas para a execução das atividades de campo e sistematização das informações obtidas, com vistas a facilitar sua interpretação e uso. O principal paradigma tem sido a caracterização das relações solo-paisagem (Hudson, 1992, Bui & Moran, 2001). Os produtos típicos de um levantamento convencional de solos são um relatório com informações descritivas sobre a ocorrência, características e propriedades das classes de solo identificadas, e um mapa com a representação cartográfica da sua distribuição espacial (Kerry et al., 2012).

Os mapas convencionais de solos são constituídos de polígonos que representam unidades de mapeamento, cada qual delineando a distribuição de um ou mais solos na paisagem. O delineamento de cada unidade de mapeamento é uma predição empírica da ocorrência dos solos com base em um modelo conceitual solo-paisagem desenvolvido a partir de observações esparsas. Uma unidade de mapeamento é denominada de simples quando representa uma única classe de solo, e de combinada, quando representa um arranjo de duas ou mais classes. Os solos em uma unidade de mapeamento são denominados componentes, e suas proporções são aproximadamente definidas (Bui, 2004; Stum et al., 2010; Ashtekar & Owens, 2014; Odgers et al., 2014b).

Atualmente os mapas convencionais são a fonte de informação mais comum sobre a distribuição dos solos, mas eles não suprem uma série de novas demandas, notadamente as voltadas à modelagem e aplicações ambientais (Hartemink et al., 2010; Häring et al., 2012; Miller, 2012; Omuto et

al., 2013). Entre os fatores que limitam seu uso está a grande heterogeneidade nos métodos e sistemas de classificação dos solos e nas escalas e áreas de cobertura dos mapeamentos, e sua qualidade indeterminada. Além disso, os mapas convencionais são constituídos de polígonos discretos nos quais a variação espacial dos solos e de suas propriedades é desconhecida, enquanto na natureza essa variação é contínua e pode ocorrer em curtas distâncias (Heuvelink & Webster, 2001; Carré et al., 2007; Ashtekar & Owens, 2014).

A maior parte das demandas atuais necessita de conhecimento mais detalhado sobre a distribuição espacial dos solos e de seus atributos (Häring et al., 2012; Omuto et al., 2013). Entretanto, a dificuldade para realizar novos levantamentos torna necessário extrair o máximo possível dos mapas convencionais existentes. Uma alternativa é reestruturá-los para satisfazer as necessidades das demandas, e uma das abordagens possíveis é a desagregação espacial das unidades de mapeamento. Basicamente, a desagregação consiste em usar técnicas de mapeamento digital de solos (MDS) e covariáveis ambientais disponíveis para descrever a variação espacial que não está explícita nos mapas convencionais. O objetivo é produzir novos mapas *raster* com a alocação espacial das classes ou propriedades dos solos que compõem as unidades de mapeamento (Nauman et al., 2014; Subburayalu et al., 2014).

No caso das classes, a desagregação pode ser realizada tanto para individualizá-las dentro das unidades de mapeamento sem desfazer suas bordas (Bui & Moran, 2001; Thompson et al., 2010; Häring et al., 2012) quanto para recriar mapas de solos sem os limites dos polígonos originais (de Bruin et al., 1999; Bui & Moran, 2003; Hansen et al., 2009; Wei et al., 2010; Smith et al., 2012; Nauman & Thompson, 2014; Subburayalu et al., 2014; Odgers et al., 2014a). Alguns estudos também demonstraram a possibilidade de produzir novos mapas que descrevam a variação espacial contínua de algumas das propriedades dos solos que compõem as unidades de mapeamento (Kerry et al., 2012; Nauman et al., 2012; Odgers et al., 2014b; Malone et al., 2014).

O uso de árvores de decisão tem se mostrado especialmente útil e bem sucedido na desagregação de mapas convencionais de solos (Nauman et al., 2014; Odgers et al., 2014b). Árvores de decisão constituem uma técnica não paramétrica que agrupa os dados por meio de uma série de subdivisões

binárias sucessivas no espaço de covariáveis, a raiz, formando um conjunto hierárquico de regras de decisão. Cada subdivisão é denominada de nó e o nó terminal constitui uma folha, um conjunto homogêneo segundo as regras de decisão. O controle do tamanho da árvore é chamado de poda e busca evitar a ramificação excessiva reagrupando recursivamente folhas pequenas em outros subconjuntos homogêneos. A homogeneidade de um subconjunto pode ser definida por uma série de medidas, as quais constituem a principal diferença entre os diversos algoritmos (Bui & Moran, 2003; Scull et al., 2005). Árvores de decisão são capazes de processar grandes volumes de covariáveis contínuas e discretas, independem de uma distribuição normal e podem facilmente caracterizar relações não lineares. A categorização hierárquica produz regras de decisão facilmente interpretáveis que se adequam ao paradigma solo-paisagem, pois mimetizam o processo mental usado na elaboração do mapa convencional de solos (Bui & Moran, 2001; Elnagar & Noler, 2010).

Independente do método ou técnica de mapeamento preditivo, as abordagens podem ser agrupadas em basicamente dois tipos: baseada em dados ou baseada em conhecimento (Smith et al., 2012; Ashtekar & Owens, 2014). A primeira usa modelagem empírica para estabelecer quantitativamente as relações entre a variável de interesse (classe ou propriedade) e as covariáveis preditoras com base em um conjunto de dados observados, para depois efetuar a predição. O ajuste dos modelos usa princípios estatísticos e o nível de confiança depende da disponibilidade de dados observados (perfis) em quantidade e com distribuição apropriada (Hansen et al., 2009). A segunda abordagem substitui o ajuste estatístico pelo conhecimento formal existente sobre os solos de uma área, sintetizado nas descrições em relatórios de levantamentos ou obtido de julgamentos feitos por especialistas. Ela se beneficia do fato de que os mapas convencionais de solos são representações do conhecimento estruturado sobre a distribuição dos solos na paisagem (Bui, 2004). É uma abordagem especialmente útil quando há poucos dados observados e, devido ao conhecimento subjacente, costuma render níveis de confiança superiores à abordagem meramente estatística (Wei et al., 2010; Ashtekar & Owens, 2014; Odgers et al., 2014a).

Os objetivos deste estudo foram propor, aplicar e validar procedimentos para desagregação de unidades de mapeamento combinadas,

com vistas a individualizar as classes de solo que as compõem para formar unidades simples ou ao menos mais uniformes quanto às relações solo-paisagem. A premissa é que as informações analíticas e descritivas do relatório de levantamento de solos podem ser traduzidas em um modelo conceitual que, aplicado a um conjunto de covariáveis ambientais relevantes e com resolução espacial apropriada, possibilita obter uma versão melhorada do mapa de solos existente.

4.2. Material e métodos

4.2.1. Área de estudo

A área de estudo situa-se na zona de produção vitivinícola Serra Gaúcha, localizada na porção nordeste do estado do Rio Grande do Sul. Compreende um retângulo da articulação para o mapeamento sistemático na escala 1:50.000 correspondente à folha SH.22-V-D-II-2 (Bento Gonçalves). Sua superfície totaliza 67.372 hectares de área mapeada, entre as latitudes 29°15'S e 29°00'S e as longitudes 51°45'W e 51°30'W (Figura 3).

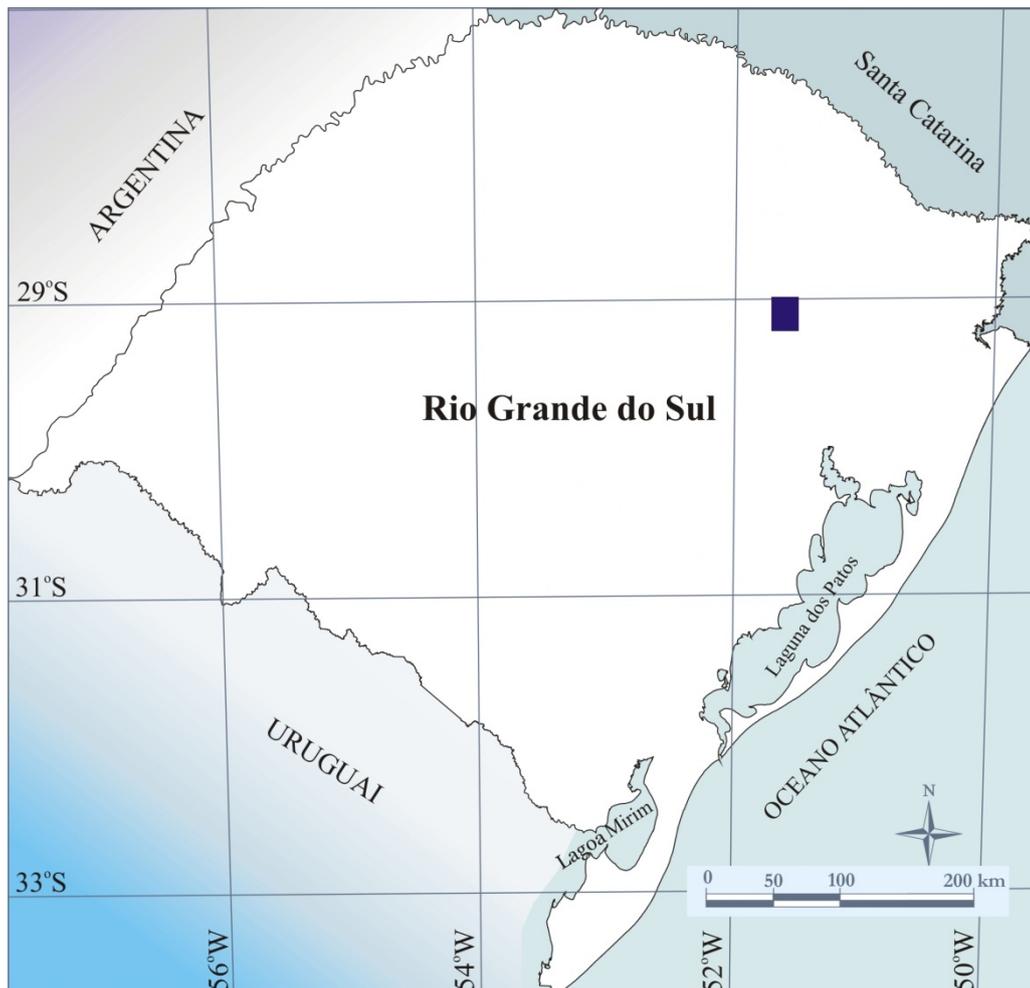


Figura 3. Localização da folha Bento Gonçalves no Rio Grande do Sul.

Segundo a classificação de Köppen, o tipo climático predominante na área de estudo é o Cfb, subtropical úmido com verões amenos, característico das áreas de maior altitude, encontrando-se em menor extensão também o tipo Cfa, subtropical úmido com verões quentes, nas áreas mais baixas. A precipitação pluvial anual totaliza 1.736 mm (Embrapa, 2008), com o menor acumulado em maio e o maior em setembro, porém a amplitude da variação anual não caracteriza um período chuvoso e outro seco. A temperatura, por outro lado, apresenta sazonalidade clara, com verão e inverno bem definidos, sendo janeiro o mês mais quente e junho o mais frio.

A geologia corresponde à Formação Serra Geral, sendo que a maior parte da área situa-se sobre a sequência superior (fácies Caxias), constituída de rochas efusivas ácidas, com uma porção menor sobre a sequência inferior (fácies Gramado), de composição predominantemente básica. A área de estudo insere-se na região geomorfológica Planalto das Araucárias e na região fisiográfica da Encosta Superior do Nordeste (IBGE, 1986). De maneira geral, o relevo é complexo, com altitudes entre 40 e 740 m e grandes variações de declividade e de exposição solar. A topografia é ondulada no topo, com bordas escarpadas e recortadas e drenagem fortemente encaixada, em cujas encostas situam-se as áreas mais íngremes.

Em função dos diferentes materiais de origem e do relevo movimentado, a distribuição dos solos é altamente heterogênea. De forma geral, nas áreas mais baixas ocorrem solos de média ou alta fertilidade, nas porções intermediárias do relevo predominam solos rasos e pedregosos, e nas áreas mais altas os solos tendem a ter maiores teores de matéria orgânica e de alumínio trocável (Flores et al., 2012). A floresta original foi intensamente modificada a partir de 1874, com a colonização italiana. Atualmente a estrutura fundiária é de pequenas propriedades rurais, cuja principal atividade é a vitivinicultura, e os remanescentes da floresta nativa restringem-se às áreas mais acidentadas e isoladas (IBGE, 1986).

4.2.2. Material utilizado

O material utilizado neste estudo compreendeu parte de um levantamento semidetalhado de solos (Flores et al., 2007c), incluindo o respectivo relatório de levantamento e um mapa convencional com escala de apresentação 1:50.000. Os resultados do capítulo anterior conferem a este

mapa potencial para ser aprimorado (Sarmiento et al., 2014). Também foi utilizada uma base cartográfica digital na escala 1:50.000, um modelo numérico do terreno (MNT) e uma imagem do satélite Landsat 5 TM, os três extraídos de uma base de dados disponível para a região (Hasenack & Weber, 2007). Adicionalmente, para a validação dos mapas desagregados, utilizou-se um conjunto de *pedons* independentes disponíveis em parte da área de estudo (Flores et al., 2012). Todos os dados geoespaciais foram estruturados na projeção UTM, fuso 22, referenciadas ao *datum* SAD69, com resolução espacial de 30 m. O processamento e análise foi executado com os programas ArcGIS (ESRI, 2013) e Idrisi (Eastman, 2012).

4.2.3. Metodologia

A metodologia utilizada inspira-se nos procedimentos empregados por Thompson et al. (2010), Nauman et al. (2012) e Nauman & Thompson (2014), com adequações para a escassez de dados descritivos da realidade local. A definição de regras a partir de bases de dados sistematizadas foi substituída por decisões baseadas em conhecimento especialista, com grau moderado de intervenção manual, contando-se com a colaboração de um dos pedólogos responsáveis pelo levantamento de solos original. O estudo concentrou esforços na adaptação, desenvolvimento e teste de uma metodologia de desagregação em detrimento da extensão da área contemplada. A área de estudo selecionada é representativa das condições de paisagem e de solos da região, bem como da crescente demanda local por informações mais detalhadas sobre os mesmos (Flores et al, 2012). Uma visão geral da metodologia utilizada para a desagregação das unidades de mapeamento de solos na área de estudo é apresentada na Figura 4.

Inicialmente, procedeu-se a uma leitura cuidadosa do mapa de solos e do relatório de levantamento, com o objetivo de recuperar e sistematizar o conhecimento existente. Analisou-se o número total de polígonos no mapa, o número de unidades de mapeamento (UM) simples e combinadas, as classes de solo identificadas, a fração ocupada por classe de solo individual e as possíveis inclusões em cada UM, as principais características e propriedades dos solos, a localização e dados analíticos e descritivos dos perfis representativos, a descrição da ocorrência dos solos na paisagem, o uso atual/cobertura vegetal predominante, bem como outros aspectos relevantes.

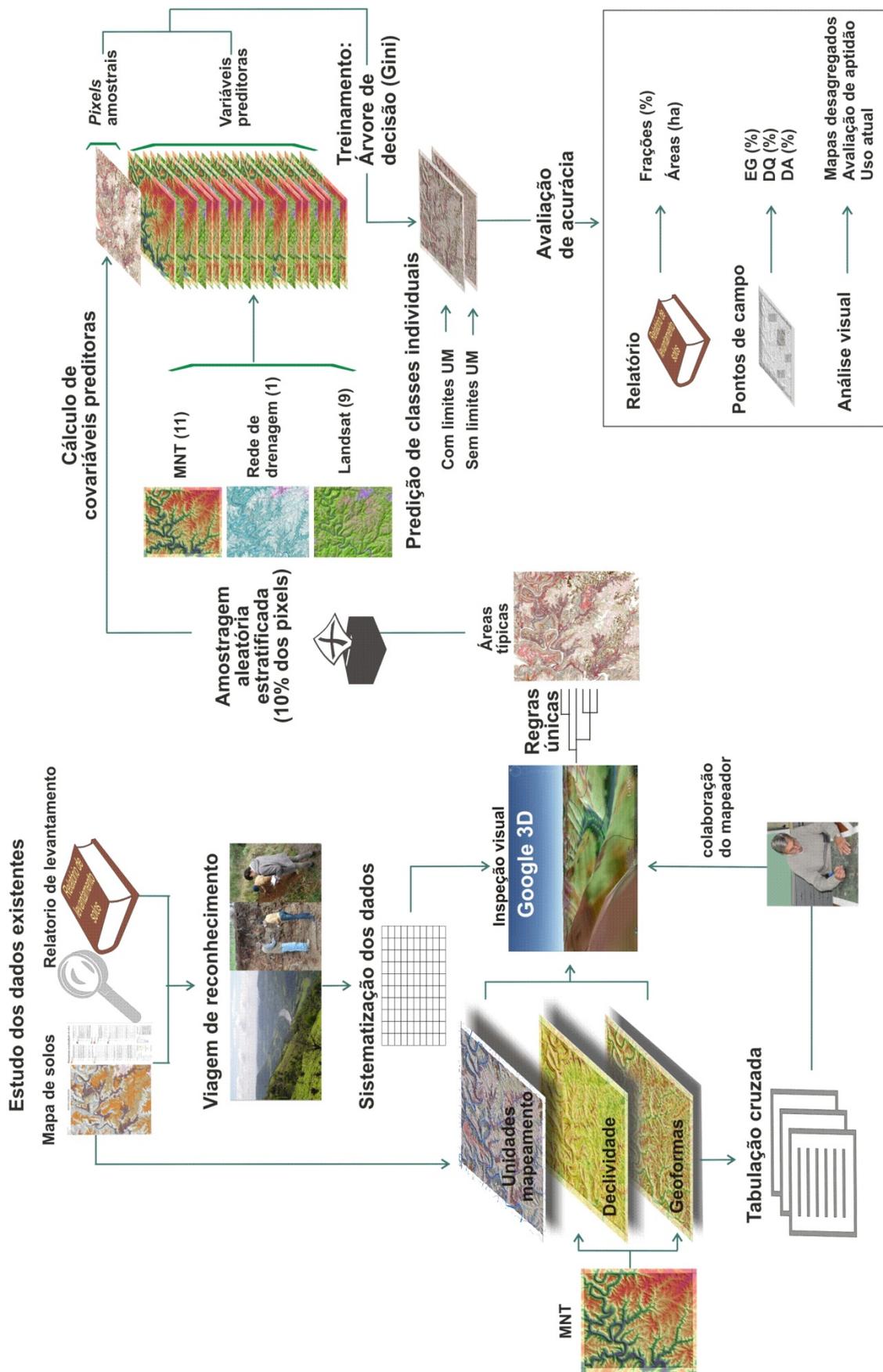


Figura 4. Fluxograma da metodologia utilizada na desagregação de unidades de mapeamento de solos na Serra Gaúcha.

Constatou-se que o mapa de solos da área de estudo é constituído de 163 polígonos que representam a distribuição espacial de 27 UM, sendo quatro UM simples e 23 UM combinadas. Destas, 14 UM delimitam duas classes e nove UM delimitam três classes de solos, todas na forma de associações de solos. A superfície das UM combinadas corresponde a 97,3% do total da área mapeada. Tomando-se a descrição completa, a qual inclui as características ou propriedades que diferenciam os solos após o quarto nível categórico, identificaram-se 42 tipos de solos distintos. Considerando-se a classificação taxonômica até o 4º nível categórico, foram identificadas 18 subgrupos pertencentes a 17 grandes grupos, 9 subordens e 5 ordens. Além dos tipos de solos, em uma das UM combinadas foi identificado ainda um afloramento rochoso.

O relatório de levantamento apresenta uma descrição detalhada das classes de solo identificadas, com ênfase nas suas características químicas, morfológicas e granulométricas, classificação sistemática e considerações sobre seu uso agrícola. No entanto, há lacunas na descrição da sua ocorrência na paisagem e na relação com as superfícies geomórficas, resumindo-se na maior parte dos casos à fase de relevo. A fração ocupada por classe de solo e a identificação de possíveis inclusões em cada UM também apresenta inconsistências. Por exemplo, das 23 UM combinadas, em cinco não consta a fração ocupada pelas suas classes individuais, e em somente quatro foram identificadas as classes que podem ocorrer como inclusões. Das quatro UM simples, em apenas uma consta o nível de pureza estimado, embora em todas tenham sido identificadas as possíveis inclusões.

O detalhamento vago em alguns itens no relatório de levantamento tornou necessária a realização de expedições em campo para complementar o conhecimento existente, principalmente quanto às relações solo-paisagem. A área de estudo foi percorrida para observação da distribuição dos solos e registro da posição na paisagem, associação com formas do relevo, uso atual e cobertura vegetal, além de outros aspectos úteis para identificar a ocorrência típica de cada classe de solo, tanto em UM combinadas quanto em UM simples. Foram visitados 15 pontos na área de estudo e nas imediações, selecionados em função da representatividade para as observações a serem complementadas e facilidade de acesso. Em cada local visitado, foram também

registradas as coordenadas com GPS e adquiridas fotografias da paisagem em várias direções, bem como anotadas as informações descritivas pertinentes. As informações obtidas em campo e aquelas extraídas do relatório do levantamento foram depois sistematizadas em tabelas com vistas a facilitar a visualização e interpretação (Wei et al., 2010, Smith et al., 2012). Em certa medida, o conteúdo dessas tabelas organizadas manualmente reproduz a informação descritiva extraída diretamente das tabelas de atributos do SSURGO (*Soil Survey Geographic Database*) por Thompson et al. (2010), Nauman et al. (2012) e Nauman et al. (2014).

A partir das informações descritivas sistematizadas, buscou-se identificar e individualizar áreas de ocorrência típicas, que fossem mais representativas das classes de solo que compõem cada UM, semelhante ao efetuado por Hansen et al. (2009) e Nauman & Thompson (2014). Nesta fase, utilizou-se como suporte para a representação espacial um mapa com classes de declividade, para representar as fases de relevo (IBGE, 2007), e outro com classes de geoformas (MacMillan, 2003), para caracterizar as diferentes formas do terreno. Em contraste com a maioria das medidas morfométricas computadas para *pixels* individuais, as classes de geoformas constituem áreas contíguas que se aproximam do conceito de superfícies geomórficas.

Os limites das UM e os mapas de classes de geoformas e de declividades foram submetidos a uma tabulação cruzada para identificar em cada UM as combinações existentes entre suas classes. Em seguida os limites das UM e os mapas de declividades e geoformas foram convertidos para o formato *Keyhole Markup Language* (KML) e carregados no programa Google Earth (Google, 2014), buscando-se criar visões 3D com pontos de vista similares aos fotografados *in loco* (Figura 5). Para cada UM, conduziu-se uma inspeção visual para analisar e interpretar uso/cobertura do solo, declividades e geoformas frente ao conhecimento estruturado a partir das informações descritivas do relatório e das observações em campo. As combinações de declividades e geoformas mais características de cada classe de solo individual por UM foram então selecionadas na tabulação cruzada (Figura 6). Integrando forma e posição relativa no relevo, declividade e uso e cobertura do solo, essas combinações representam concretamente um modelo solo-paisagem para a área de estudo (Barringer et al., 2008).

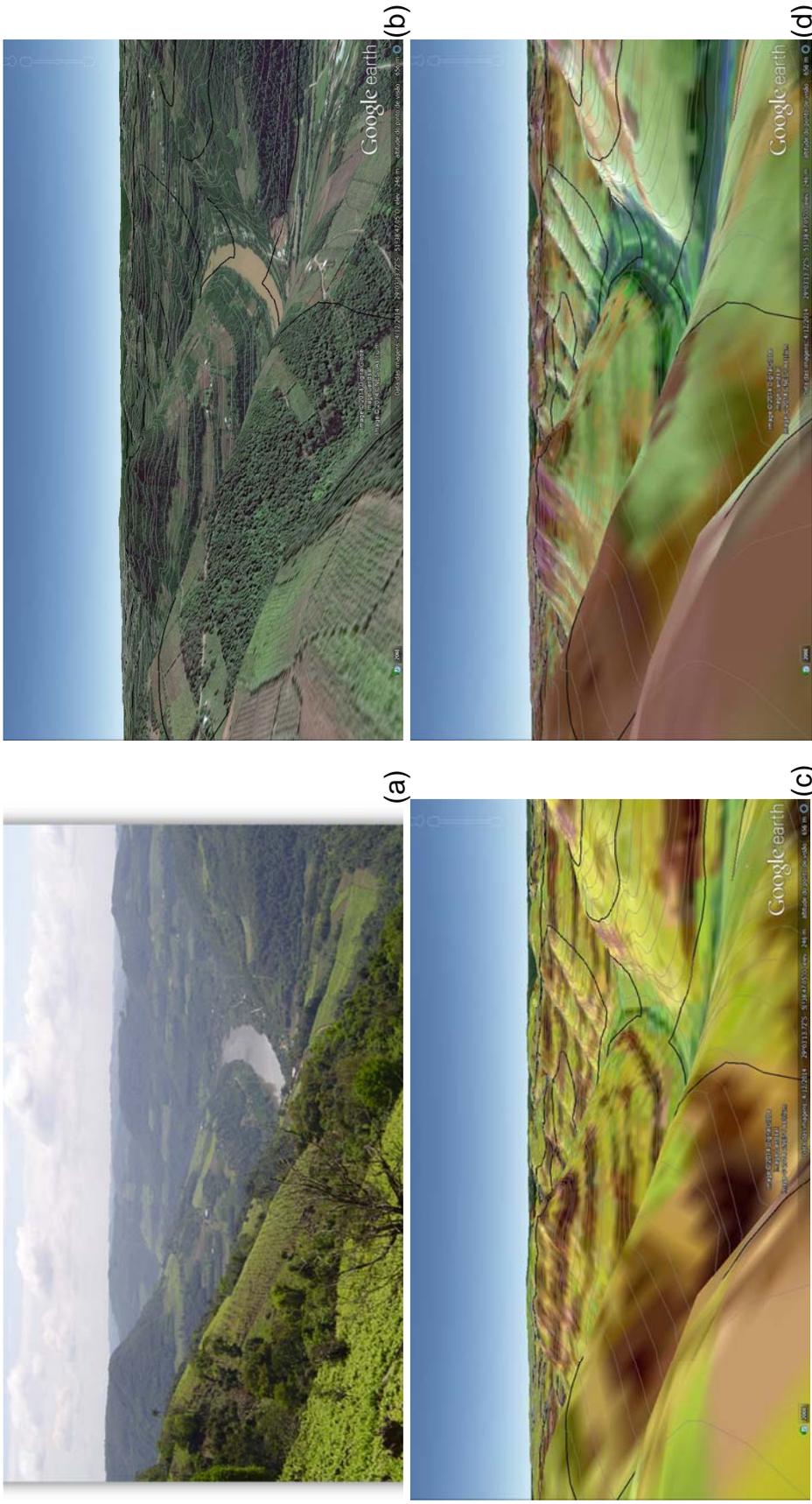


Figura 5. (a) Fotografia da paisagem adquirida em um local visitado em campo; (b) ponto de vista aproximadamente igual no Google Earth, com sobreposição dos limites das UM (linhas pretas) às imagens de satélite; (c) às classes de declividade; e (d) às classes de geoformas.

A partir das combinações selecionadas, foram criadas regras para a individualização das áreas de ocorrência típicas de cada classe de solo nas 27 UM, totalizando 59 regras únicas (Tabela 4). O número de regras foi maior que o de solos porque um determinado solo pode ser componente de mais de uma UM e, em cada uma delas, pode ocorrer em situações distintas na paisagem. Assim, em alguns casos foi necessário criar mais de uma regra para individualizar *pixels* típicos do mesmo solo. O conjunto de regras únicas foi posteriormente traduzido para expressões lógicas em SIG a fim de produzir, a partir das UM, das declividades e das geoformas, um novo mapa somente com os *pixels* representativos de cada solo individual. Em média, a superfície coberta por esses *pixels* típicos totalizou cerca de 60% a 65% da área de cada UM delineada no mapa de solos original.

Geoformas	Declividade						Fração (%)
	0-3%	3-8%	8-20%	20-45%	45-75%	>75%	
LCR	0.02	0.30	1.36	0.17	0.01	0.00	1.87
DSH	0.91	6.12	23.48	12.47	1.32	0.07	44.37
UDE	0.03	0.34	1.77	1.45	0.17	0.01	3.77
BSL	0.05	2.41	9.76	4.05	0.35	0.02	16.64
DBS	0.09	1.31	6.60	4.64	0.56	0.07	13.27
CBS	0.11	1.51	7.38	4.26	0.50	0.03	13.79
TER	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SAD	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MDE	0.02	0.00	0.00	0.00	0.00	0.00	0.02
FSL	0.21	1.20	2.50	0.95	0.07	0.01	4.95
TSL	0.01	0.24	0.22	0.00	0.00	0.00	0.47
FAN	0.00	0.22	0.42	0.06	0.00	0.00	0.70
LSM	0.00	0.02	0.05	0.04	0.00	0.00	0.11
LLS	0.01	0.00	0.00	0.00	0.00	0.00	0.01
DEP	0.02	0.00	0.00	0.00	0.00	0.00	0.02
Fração (%)	1.50	13.69	53.55	28.08	2.98	0.20	100.00
	C1		C2		C3		
	50.74		17.11		7.23		

Geoformas	Declividade						Fração (%)
	0-3%	3-8%	8-20%	20-45%	45-75%	>75%	
LCR	0.00	0.00	0.18	0.00	0.00	0.00	0.18
DSH	0.11	0.25	1.57	0.18	0.06	0.00	2.17
UDE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BSL	0.00	0.61	3.28	5.29	1.40	0.11	10.70
DBS	0.07	0.43	2.31	6.41	2.00	0.18	11.40
CBS	0.00	0.06	0.50	1.63	0.63	0.02	2.83
TER	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SAD	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MDE	0.01	0.00	0.00	0.00	0.00	0.00	0.01
FSL	1.55	5.01	20.07	12.22	1.46	0.07	40.37
TSL	0.70	4.13	8.43	1.56	0.03	0.00	14.85
FAN	0.14	1.58	6.23	3.98	0.36	0.01	12.30
LSM	0.12	0.59	2.02	1.58	0.01	0.00	4.32
LLS	0.19	0.03	0.00	0.00	0.00	0.00	0.23
DEP	0.56	0.06	0.02	0.00	0.00	0.00	0.64
Fração (%)	3.45	12.74	44.61	32.85	5.95	0.39	100.00
	C1						
	65.59						

(a)

(b)

Figura 6. Exemplos da tabulação cruzada de classes de declividade e de geoformas, com a identificação da ocorrência típica das classes de solos em: (a) uma UM combinada com três classes; (b) uma UM simples. As letras e números abaixo dos quadros indicam as componentes e as respectivas frações na UM (LCR: level crest; DSH: divergent shoulder; UDE: upper depression; BSL: back slope; DBS: divergent back slope; CBS: convergent back slope; TER: terrace; SAD: saddle; MDE: mid slope depression; FSL: foot slope; TSL: toe slope; FAN: lower slope fan; LSM: lower slope mound; LLS: level lower slope; DEP: depression).

A individualização com base em conhecimento local confere a esses *pixels* grau de confiança e pureza elevado em relação à classe considerada, características desejáveis para MDS (Odgers et al., 2014a). No entanto, o número de *pixels* típicos resultou excessivamente grande para viabilizar o processamento, além de representar risco de sobreposição. Dessa forma,

optou-se por selecionar um subconjunto amostral de modo a equilibrar a viabilidade de processamento e a caracterização das classes de solo em termos de covariáveis. A partir de considerações feitas em outros estudos, o subconjunto amostral foi dimensionado em 10% do total (Zhu, 2000; Bui e Moran, 2003; Grinand et al., 2008). A seleção foi efetuada por meio de uma amostragem aleatória estratificada (Finnen & Menza, 2013), obtendo-se ao final cerca de 50.000 *pixels* típicos.

Tabela 4. Exemplos de regras para individualização da ocorrência típica das classes de solo em três UM combinadas e em uma UM simples, a partir de classes de declividade e de geoformas (LCR: level crest; DSH: divergent shoulder; BSL: back slope; DBS: divergent back slope; CBS: convergent back slope).

UM	Declividade	Geoformas	Componente da UM
PACd 1	<20%	DSH	ARGISSOLO ACINZENTADO Distrófico típico A proeminente textura muito argilosa relevo suave ondulado
	8-20%	DBS,CBS	CAMBISSOLO HÁPLICO Ta Distrófico típico A proeminente textura argilosa relevo ondulado
	20-45%	DSH, DBS, CBS	
	20-45%	BSL	NEOSSOLO LITÓLICO Eutrófico típico textura média fase pedregosa relevo ondulado
PVd 1	>45%	Qualquer	ARGISSOLO VERMELHO Tb Distrófico típico A moderado textura média/muito argilosa relevo forte ondulado
	<20%	Qualquer	CHERNOSSOLO HÁPLICO Ta Órtico típico textura média relevo forte ondulado
	20-45%	DSH	
PVd 2	20-45%	<> DSH	ARGISSOLO VERMELHO Tb Distrófico típico A moderado textura média/muito argilosa relevo ondulado
	<20%	Qualquer	NEOSSOLO LITÓLICO Eutrófico típico A moderado textura média fase pedregosa relevo ondulado
	20-45%	DSH	
	20-45%	BSL	
CHd 1	45-75%	Qualquer	
	<20%	LCR, DSH, BSL	CAMBISSOLO HÚMICO Distrófico típico textura argilosa relevo suave ondulado

Os *pixels* amostrais foram posteriormente usados como áreas de treinamento em um algoritmo de árvore de decisão. O algoritmo utilizado emprega o índice de impureza de Gini para avaliar a qualidade das subdivisões durante a construção da árvore. A impureza de Gini é uma medida da heterogeneidade interna dos *pixels* da classe formada em um determinado nó. Durante o processo de construção da árvore, ela é minimizada com vistas a criar folhas com somente uma classe, ou com um índice de impureza próximo de zero (Nauman & Thompson, 2014; Nauman et al., 2014). No processo de treinamento foi utilizado um conjunto de 21 covariáveis ambientais explanatórias, consideradas potencialmente úteis para caracterizar os fatores

de formação dos solos e as relações solo-paisagem em diversas situações (Hansen et al., 2009; Böttinger et al., 2010; Stum et al., 2010; Nauman et al., 2014, Odgers et al., 2014a) (Tabela 5).

Tabela 5. Covariáveis explanatórias e fatores de formação dos solos representados pelas mesmas na Serra Gaúcha (C: clima, R: relevo, P: material de origem, O: organismos, T: tempo).

Covariável preditora	Fator de formação
Altitude	R, C
Declividade	R
Orientação da encosta	R, C
Curvatura mínima	R
Curvatura máxima	R
Curvatura global	R
Distância da rede de drenagem	R, P
Direção de fluxo	R
Acúmulo de fluxo	R
Índice de poder de escoamento (<i>Stream Power Index - SPI</i>)	R
Índice de umidade topográfica ¹ (<i>Topographic Wetness Index-TWI</i>)	R
Índice de posição topográfica/2 (<i>Topographic Position Index-TPI</i>)	R
Landsat NDVI	O, C
Landsat B3/B2	P
Landsat B3/B7	P
Landsat B5-B2/B5+B2	P
Landsat (B5-B7/B5+B7)	P, T
Landsat (B4-B5/B4+B5)	P, O, R
Landsat (B3-B7/B3+B7)	P, O, R
Landsat (B5-B1/B5+B1)	P, O, R
Landsat (B4-B7/B4B7)	P, O, R

¹ Computado com raio de 100 m; ² Computado com raio maior de 500 m e raio menor de 200 m.

Usualmente, a árvore de decisão necessita ser podada durante o processo de construção para evitar a ramificação excessiva e a formação de folhas com apenas poucos *pixels* (Bui & Moran, 2003; Scull et al., 2005). Todavia, em função do número de classes de solo, da complexidade da paisagem e da quantidade de covariáveis explanatórias, concluiu-se que a árvore deveria ser altamente ramificada, tornando a poda manual impraticável. Assim, optou-se por uma poda automática controlada por área (Bui & Moran, 2001), definindo-se como limite para a supressão de nós terminais uma superfície equivalente 0,01% da área total (cerca de 90 *pixels*). Dessa forma, os *pixels* correspondentes às folhas excluídas foram recursivamente alocados a outras classes de modo a minimizar a heterogeneidade interna das mesmas e os erros de predição, até estes atingirem a estabilidade (Lippitt et al., 2008).

A árvore de decisão construída na etapa de treinamento foi posteriormente aplicada para a predição da ocorrência de classes de solo individuais em toda a área de estudo. A predição consistiu em estimar uma classe de solo para cada *pixel* de modo a gerar um novo mapa de classes de solo individualizadas, ou seja, desagregadas. Esse processo foi executado de duas formas distintas. Na primeira, os limites das UM do mapa convencional de solos foram usados como restrições para forçar a alocação dos *pixels* a uma das classes originalmente descritas em cada UM. Na segunda, os limites das UM foram ignorados e cada *pixel* da área de estudo foi alocado à classe de solo mais provável com base somente nos valores locais das covariáveis, podendo ser qualquer uma das classes identificadas na legenda do mapa.

Para avaliar o resultado dos dois modos de predição, foi realizada uma comparação qualitativa dos valores descritos no relatório de levantamento para o número de tipos de solo e para a fração ocupada pelos mesmos em cada UM com os correspondentes obtidos nos mapas desagregados (Bui & Moran, 2001; Roecker et al., 2010; Odgers et al., 2014a). A quantificação desses valores para os mapas desagregados foi efetuada mediante uma intersecção de cada um com os polígonos das UM do mapa convencional. Também foi comparada a superfície total de cada classe de solo individual nos mapas desagregados e no mapa convencional, sendo que neste ela foi calculada a partir da área dos polígonos de cada UM e da fração das respectivas classes informada no relatório (Li et al., 2012).

Os mapas desagregados também foram validados com base em um conjunto de *pedons* independentes, a maioria dos quais foi coletada em um levantamento detalhado de solos existente para uma porção da área de estudo. Para a avaliação de acurácia considerou-se a concordância espacial das coordenadas desses *pedons* com a localização de *pixels* individuais nos mapas desagregados e, adicionalmente, com um raio de busca (*buffer*) de 1 *pixel* em torno de cada *pedon*. Neste caso, a predição foi contabilizada como correta quando a classe de solo predita em ao menos um dos *pixels* no raio de busca foi igual à observada no *pedon* de referência. O uso de um raio de busca teve o objetivo de detectar a transição gradual entre classes que ocorre na realidade e de compensar eventuais incertezas no referenciamento espacial dos *pedons* e nas covariáveis ambientais utilizadas (Smith et al., 2012; Nauman & Thompson,

2014). Para cada mapa desagregado quantificou-se a percentagem de acertos, denominada de exatidão global (EG), e as discordâncias de quantidade (DQ) e de alocação (DA) das classes (Pontius & Millones, 2011). Os *pedons* independentes também foram utilizados para computar a EG do mapa convencional de solos, calculando-se a percentagem de acertos das classes nas UM cujos polígonos coincidiram com a localização daqueles.

Por último, foi ainda efetuada uma análise visual dos mapas de classes de solo desagregadas em comparação com o mapa convencional original constituído de polígonos. Além da distribuição espacial das classes propriamente ditas, também foram analisados os resultados de uma avaliação de aptidão dos solos para a viticultura efetuada sobre as classes desagregadas em comparação com uma avaliação previamente realizada sobre as UM do mapa convencional, usando os mesmo critérios (Flores et al., 2007d).

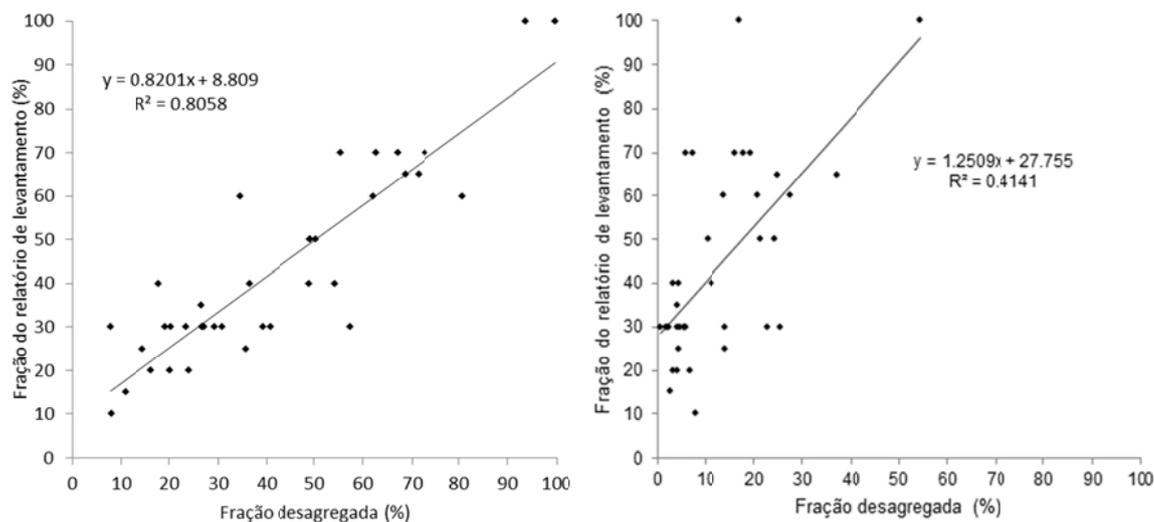
4.3. Resultados e discussão

4.3.1. Fração, número e área das classes desagregadas

A relação entre a fração desagregada e a fração da área informada no relatório para as classes em cada UM aparentemente foi mais satisfatória no procedimento em que os limites das UM foram usados como restrição na predição (Figura 7a). Considerando-se todas as UM, o erro absoluto médio, ou seja, a média das diferenças entre a fração desagregada e a fração de cada classe informada no relatório de levantamento, foi de -2,8%. O número de classes de solo individualizadas em cada UM neste procedimento não diferiu do número informado no relatório devido à restrição à predição de outras classes imposta pelos limites das UM no processo de desagregação.

No procedimento de desagregação em que os limites das UM foram desconsiderados, os resultados sugerem uma tendência de subestimativa da fração desagregada em relação à fração informada no relatório (Figura 7b), com erro absoluto médio de -33%. De fato, nesse mapa desagregado constatou-se a predição de classes de solo que não estavam listadas no relatório de levantamento para as UM, o que reduziu a proporção das que estavam descritas. Isso não significa um problema; ao contrário, tais classes podem ser inclusões não mapeadas originalmente devido a dificuldades logísticas ou que não foram identificadas no levantamento de solos original (Subburayalu et al., 2014). Como consequência da predição de novas classes,

o número de classes individualizadas por UM foi maior que o número informado no relatório de levantamento, chegando a um caso extremo de 10 vezes.



(a) (b)
 Figura 7. Relação entre a fração das componentes de cada UM após a desagregação e a fração declarada no relatório de levantamento: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.

Em ambos os procedimentos de desagregação, os desvios na fração desagregada em relação à descrita no relatório foram inversamente proporcionais à extensão das classes de solo individuais por UM (Figura 8). Esse resultado é positivo do ponto de vista do uso prático dos mapas desagregados, pois os maiores desvios corresponderão a uma parcela reduzida da área total mapeada, portanto de baixo risco para a tomada de decisão (Rossiter, 2004). No caso da desagregação restrita pelos limites das UM, os erros apresentaram distribuição aproximadamente simétrica em torno de zero (Figura 8a) porque a superestimativa de uma classe dentro de determinada UM causou a subestimativa das demais em proporção equivalente. Na desagregação livre os desvios foram sempre negativos (Figura 8b) devido à predição de classes não descritas no relatório de levantamento para as UM.

Os dois únicos casos com desvios grandes associados a áreas também grandes, maiores que 5.000 ha, foram verificados na maior UM combinada do mapa convencional de solos. Ela ocupa 35,73% da área total mapeada e é composta de três classes, sendo que a dominante ocupa 60% da superfície, a segunda 30% e a terceira 10%, segundo o relatório de

levantamento. Considerando-se a classificação taxonômica até o quarto nível do SiBCS, a classe dominante é um Neossolo Litólico Eutrófico típico, a segunda componente um Cambissolo Háplico Ta Eutrófico típico e a terceira um afloramento rochoso. Na desagregação restrita pelos limites das UM, a fração da primeira resultou menor que a da segunda, inversão que é evidenciada pelos dois pontos com maior desvio negativo e positivo na Figura 8a. Na desagregação livre a classe dominante permaneceu com a maior proporção, mas o desvio negativo em relação à sua fração informada no relatório foi maior que na segunda componente, conforme ilustra o ponto mais à direita na Figura 8b.

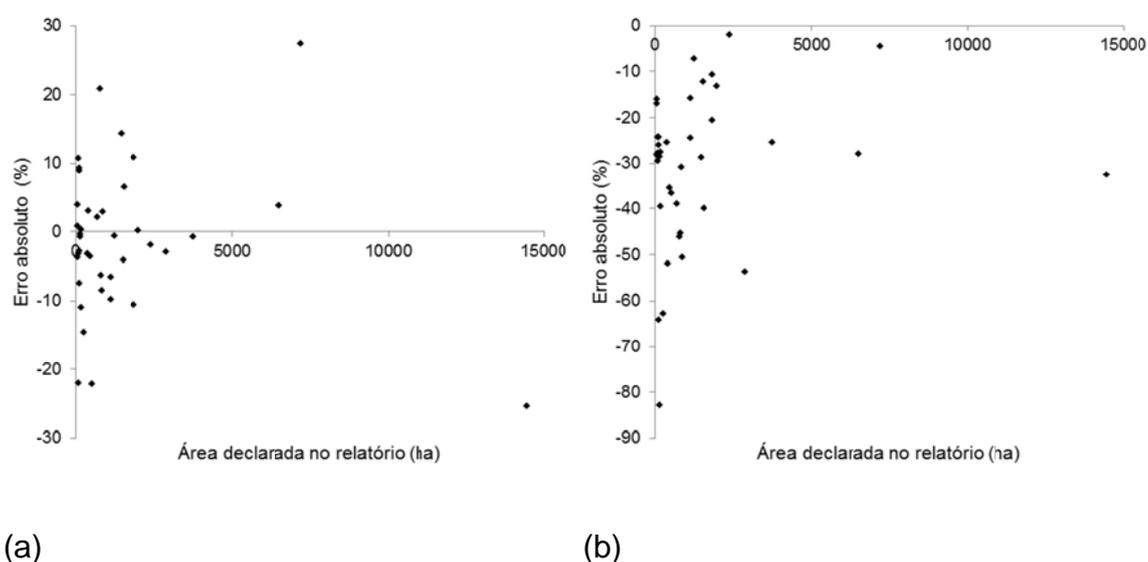


Figura 8. Diferença entre a fração das componentes de cada UM após a desagregação e sua fração declarada no relatório (erro absoluto) em relação à área das classes de solo na UM: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.

O fato de a classe dominante dessa UM resultar consideravelmente menor em ambos os procedimentos de desagregação sugere que as proporções definidas no levantamento de solos podem não corresponder à realidade. Nesse sentido, MacMillan et al. (2010) mostraram que estimativas de proporções feitas em campo por especialistas diferentes podem divergir bastante para o mesmo local em função da subjetividade. As duas classes em questão possuem características semelhantes e são difíceis de diferenciar em campo, mesmo por pedólogos experientes, o que torna plausível a hipótese de algum equívoco. Além disso, a dificuldade de acesso a certas porções dessa UM devido ao relevo acidentado e à cobertura vegetal florestal pode ter contribuído para uma definição incerta das proporções. Por outro lado, a

incerteza nas proporções informadas é irrelevante para o uso prático dos mapas desagregados já que a distância taxonômica entre as classes em questão é pequena, pois ambas apresentam solos rasos e pedregosos (Phillips, 2013).

Considerando-se a área total mapeada, os mapas obtidos com ambos os procedimentos de desagregação praticamente se equivalem em termos da soma das áreas das suas classes individuais em relação às do mapa convencional, com R^2 acima de 0,72 (Figura 9). Todavia, a inversão nas frações das classes da maior UM anteriormente mencionada continua evidente na desagregação restrita pelos limites das UM, como ilustram os dois pontos mais afastados da reta na Figura 9a. No procedimento de desagregação em que as UM não foram consideradas a dispersão dos valores das áreas totais foi menor porque classes que tiveram *pixels* omitidos em algumas UM puderam ser preditas em outras UM nas quais não haviam sido descritas, e vice-versa (Figura 9b). No mesmo tipo de avaliação, Li et al. (2012) obtiveram R^2 entre 0,87 e 0,99 para a relação entre as áreas totais de classes desagregadas e as informadas em relatórios do CLI (Canada Land Inventory). No entanto, foram consideradas somente as áreas das classes dominantes em cada UM e o objetivo foi prever a ocorrência de classes de aptidão agrícola, que podem ser iguais para várias classes taxonômicas.

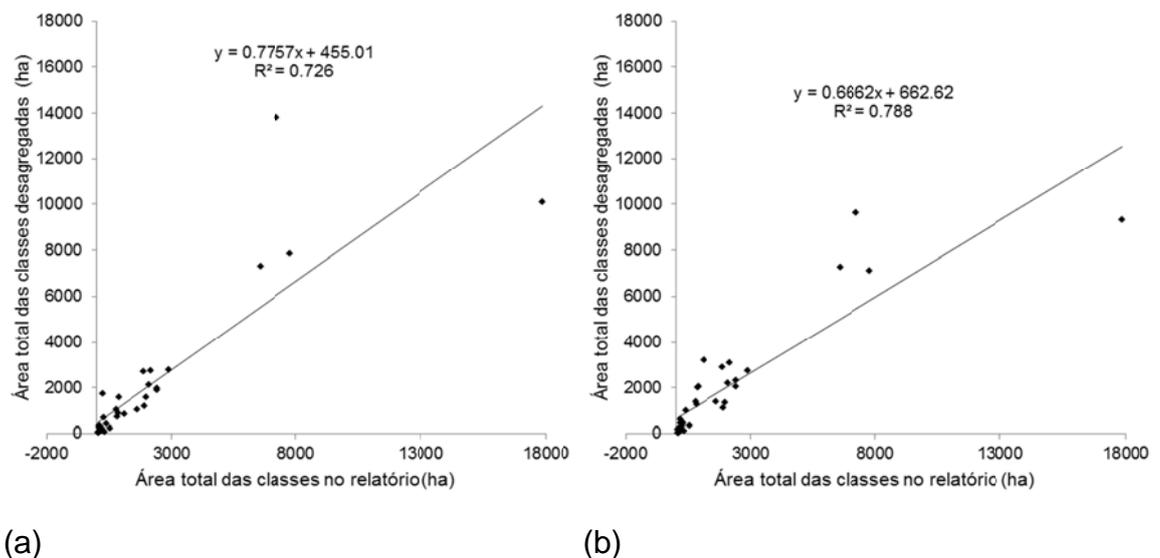


Figura 9. Relação entre a área total das classes de solo desagregadas e a área total calculada com base no relatório de levantamento: (a) restringindo a predição das classes pelo limite de cada UM; (b) sem restringir a predição pelo limite de cada UM.

As frações e áreas das classes de solo desagregadas em comparação com valores informados em relatórios de levantamento fornecem algum indicativo sobre o desempenho do processo de desagregação. Embora eles não sejam um dado de referência ideal, já que são estimativas feitas subjetivamente no levantamento de solos, certo grau de coincidência entre ambas pode significar que a desagregação foi bem sucedida ou ao menos consistente (Bui & Moran, 2001). Diferenças observadas não são erros propriamente ditos, são meramente desvios de uma estimativa em relação a outra, pois nenhuma das duas deriva de medidas efetivas em campo. Trata-se de tentar extrair o máximo da informação existente, pois os valores informados nos relatórios frequentemente são a única referência disponível sobre a distribuição espacial das classes em mapas convencionais de solos (Odgers et al., 2014a).

É provável que as frações e áreas obtidas em mapas desagregados sejam mais confiáveis que os informados em relatórios de levantamento em função da maior consistência na predição da ocorrência das classes. Uma vez estabelecido o modelo, neste caso uma árvore de decisão, ele é uniformemente aplicado a toda a área de estudo, o que não ocorre no delineamento manual das UM. Para MacMillan et al. (2010), não é realista esperar que qualquer mapa predito concorde com proporções de classes estimadas em campo por especialistas em taxas melhores que 64 a 71%, valores compatíveis com os obtidos neste estudo.

4.3.2. Validação com *pedons* independentes

A validação com *pedons* independentes fornece medidas mais concretas sobre o desempenho do processo de desagregação em relação à realidade. No entanto, a disponibilidade de *pedons* independentes costuma ser escassa, e frequentemente é necessário algum tipo de adequação a fim de aproveitar os existentes (Nauman & Thompson, 2014; Nauman et al., 2014). Neste estudo, alguns dos *pedons* disponíveis tiveram que ser descartados porque correspondiam a classes de solo não identificadas no mapa convencional sob desagregação, o que é esperado já que a intensidade dos respectivos levantamentos é diferente. Restaram 233 *pedons* aproveitáveis (Figura 10), sendo que 157 correspondem a perfis completos, com todos os dados analíticos, e 76 são pontos de observação, nos quais existe somente a

identificação das classes até o nível de subordem. A análise da legenda do mapa convencional revelou que todas as classes componentes das UM se diferenciam já no nível de subordem, de forma que não haveria prejuízo em executar a avaliação de acurácia nesse nível, podendo-se, portanto, aproveitar todos os 233 pontos.

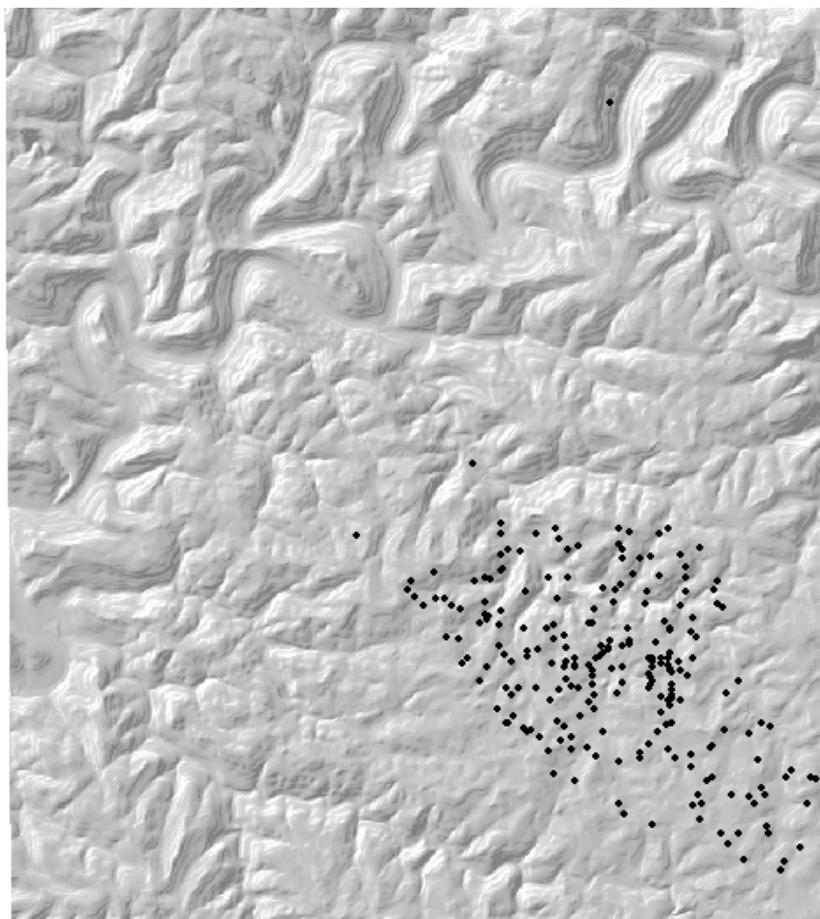


Figura 10. Localização dos 233 *pedons* independentes usados para a validação dos mapas desagregados.

A distribuição espacial dos *pedons* não é aleatória, como seria desejável para validação, e eles cobrem apenas parte da área de estudo (Figura 10), mas ainda assim representam informação valiosa para medir o quão próximas da verdade são as classes preditas em *pixels* individuais nos mapas desagregados. Além disso, a coleta de novos perfis com distribuição e número ideais em áreas extensas enfrenta dificuldades logísticas e de custo que geralmente inviabilizam a tarefa. Por essa razão, usar *pedons* existentes tem sido uma estratégia amplamente adotada (Hansen et al., 2009; Nauman et al., 2012; Nauman & Thompson, 2014; Odgers et al., 2014a; Malone et al., 2014). Quanto à representatividade, os 233 *pedons* aproveitáveis contemplam

24 dos 163 polígonos do mapa convencional de solos, 12 das 27 UM, 19 das 42 classes individuais, considerando a descrição completa, 13 dos 18 subgrupos, 12 dos 17 grandes grupos, 8 das 9 subordens e todas as 5 ordens.

Considerando-se a concordância simples dos mapas desagregados com os *pedons* independentes, a exatidão global (EG) foi de 33,9% na desagregação em que os limites das UM foram usados como restrições na predição das classes, e de 28,3% na desagregação livre (Tabela 6). A menor EG do segundo é explicada pela sua discordância de alocação (DA) mais elevada, enquanto sua discordância de quantidade (DQ) foi inferior à do primeiro. DA maior que DQ é preferível porque significa que parte de algumas classes foram preditas em *pixels* que não coincidem exatamente com a localização dos pontos de validação, mas que sua superfície se aproxima da real (Pontius & Millones, 2011). Esses valores corroboram os desvios nas frações e áreas observados para algumas classes na desagregação restrita pelos limites das UM, como evidenciado na Figura 8a e na Figura 9a.

Tabela 6. Medidas de acurácia dos mapas desagregados em relação a 233 *pedons* independentes com verdade de campo.

Forma de avaliação	Desagregação restrita			Desagregação livre		
	EG (%)	DA (%)	DQ(%)	EG(%)	DA(%)	DQ(%)
Concordância simples	33,9	28,8	37,3	28,3	36,9	34,8
Com filtro de moda	33,1	28,4	38,5	33,0	28,4	38,6
Raio de 1 <i>pixel</i>	45,5	---	---	48,5	---	---

EG: Exatidão global (percentagem de acertos); DA: Discordância de alocação; DQ: Discordância de quantidade.

O mapa obtido com o procedimento de desagregação livre apresentou maior quantidade de *pixels* de certas classes isolados entre áreas contíguas de outras, o que pode ter contribuído para sua DA mais alta, pois a chance da classe correta estar em um *pixel* ao lado do *pedon* de referência aumenta. A aplicação de um filtro de moda para reduzir a frequência de *pixels* isolados, como efetuado por Bui & Moran (2003), elevou a EG desse mapa, mas em contrapartida reduziu a do mapa obtido com a desagregação restrita pelos limites das UM. Com a filtragem, os valores de EG, DA e DQ resultaram praticamente idênticos para ambos os mapas. Quando se contabilizou os acertos para um raio de busca de 1 *pixel* em torno de cada *pedon*, a EG aumentou consideravelmente, chegando a 45,5% na desagregação restrita

pelos limites das UM e a 48,5% na desagregação livre. Neste caso não foi possível calcular DA e DQ, mas os valores de DA anteriores indicam que possivelmente a EG aumentaria ainda mais com a expansão do raio de busca. No entanto, vários *pedons* teriam que ser descartados a fim de evitar contagens duplicadas no mesmo raio devido à pequena distância entre alguns.

A percentagem de *pixels* corretamente classificados nos mapas desagregados pode ser considerada satisfatória, pois se encontra em patamar equivalente ao obtido em diversos outros estudos recentes. Subburayalu et al. (2014), por exemplo, obtiveram percentagem de acertos entre 25% e 39% nas classes desagregadas com três modelos de árvores de decisão em Ohio, usando 106 *pedons* independentes como referência. Nauman & Thompson (2014) compararam mapas desagregados por três modelos de árvores de decisão baseadas no índice de Gini, o mesmo usado neste estudo, com 87 *pedons* independentes em *West Virginia*, obtendo exatidão global entre 22% e 24% na concordância simples dos *pixels* do mapa com os *pedons* e entre 39% e 44% em um raio de busca de 60 m em torno de cada *pedon*. Odgers et al. (2014a) obtiveram EG de 22,5% em mapas desagregados com múltiplas árvores de decisão em Queensland, usando 285 *pedons* independentes como referência, e Holmes et al. (2014) obtiveram EG de 18% para as classes desagregadas com o mesmo procedimento no oeste da Austrália, em comparação com pontos independentes extraídos de uma base de dados de 48.269 perfis. A diferença destes para os exemplos anteriores é que eles buscaram prever automaticamente a ocorrência das classes individuais somente a partir das suas proporções no relatório de levantamento, sem considerar outras fontes de conhecimento.

Níveis de concordância mais elevados foram obtidos por Hansen et al. (2009), que usaram árvores de decisão e conhecimento especialista local para desagregar catenas descritas em Uganda e obtiveram EG entre 75,5% e 78,5% na comparação com 193 perfis independentes. Entretanto, o objetivo foi prever quatro classes de paisagem, as quais podem conter vários tipos de solo simultaneamente. Häring et al. (2012) obtiveram EG de 70% para classes desagregadas com árvores de decisão na Baviera em comparação com 1.812 perfis, porém aceitando somente previsões com probabilidade superior a 0,7, de forma que porção considerável da área não foi desagregada. Nauman et al.

(2014) obtiveram EG de 66,7% e 69,8% na desagregação de mapas convencionais de solos do Arizona, com dois modelos de árvores de decisão baseadas no índice de Gini, usando como referência 63 *pedons* coletados pelos próprios autores. A maioria dos pontos foi coletada ao longo de rodovias, em pontos escolhidos intencionalmente de forma a melhor representar as classes, locais onde a confusão na predição também tende a ser menor.

Quanto ao mapa convencional utilizado neste estudo, na comparação com os *pedons* independentes a percentagem de acertos nas classes das UM simples foi de 25%, embora elas teoricamente delimitem uma única classe. Nas UM combinadas a concordância foi de 33,5% para a classe dominante, de 29,6% para a segunda componente e 7,3% para a terceira. Considerando-se a concordância com qualquer uma das classes das UM a percentagem de acertos foi de 70,4%, com a ressalva de que neste caso mais de uma classe por UM foi contada como acerto. Os valores de EG obtidos para as classes individuais de solos nos mapas desagregados com os dois procedimentos são coerentes com o verificado para as classes nas UM do mapa convencional. Nesse contexto, a vantagem dos mapas desagregados é que eles indicam a localização de cada uma das classes, que nas UM do mapa convencional é desconhecida, sem reduzir a acurácia ou, inclusive, melhorando-a em relação ao original (Thompson et al., 2010; Nauman et al., 2014; Subburayalu et al., 2014).

A equivalência na EG dos dois mapas desagregados quando da aplicação de um filtro de moda, bem como o incremento na EG com o uso de um raio de busca em torno de cada *pedon*, revela algum grau de incerteza espacial (Smith et al., 2012; Nauman & Thompson, 2014). As fontes podem ser várias: as áreas típicas usadas como treinamento, as covariáveis selecionadas para caracterizar as relações solo-paisagem, o algoritmo utilizado para individualizar as classes, ou os próprios *pedons* usados como referência. Além disso, a comparação das coordenadas dos *pedons* com *pixels* discretos nos mapas desagregados, que neste estudo têm resolução espacial de 30m, favorece discordâncias porque os solos variam continuamente no espaço, e em distâncias particularmente curtas na área de estudo (Figura 11). A classe predita em um *pixel* expressa a tendência dos valores das covariáveis em uma área de 900 m², dentro da qual podem ser identificados vários solos diferentes

em campo. Eventualmente, parte do problema poderia ser atenuada com a coleta de número maior de *pedons* com distribuição mais homogênea na área de estudo, mas limitações de custo e de tempo inviabilizaram essa alternativa. A opção de aumentar a resolução espacial das covariáveis é desaconselhável, pois incrementa o ruído em relação aos dados originais de solos, que permanecem os mesmos (Roecker et al., 2010).



Figura 11. Corte de estrada na área de estudo mostrando a variação espacial dos solos ao longo de uma distância que é inferior à resolução espacial utilizada (30m).

4.3.3. Análise visual

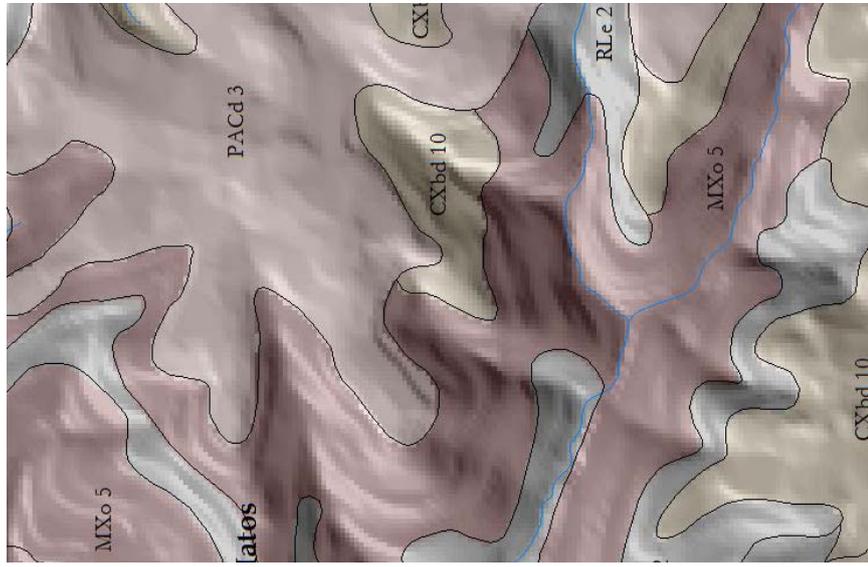
Visualmente, os mapas de classes de solo individuais resultantes de ambos os procedimentos de desagregação utilizados apresentam um incremento substancial no detalhe espacial em relação ao mapa convencional (Figura 12). Eles contêm a classe de solo individual mais provável em cada *pixel* da área de estudo, previstas com base nas suas áreas de ocorrência típicas inicialmente definidas e nos valores das covariáveis explanatórias utilizadas. Os polígonos referentes aos limites das UM no mapa convencional, que delineavam a ocorrência de uma a três classes de solos, deixaram de existir (Subburayalu et al., 2014).

As diferenças entre a aparência dos dois mapas desagregados são mais pronunciadas do que evidenciaram as frações e áreas das classes individuais e a comparação com *pedons* independentes. No mapa obtido com o procedimento em que os limites das UM foram usados como restrições na predição (Figura 12b), algumas classes desagregadas trespassaram os limites originais, aparecendo de forma contígua em UM adjacentes, enquanto outras ficaram confinadas somente à UM em que foram descritas. Devido a estas, a

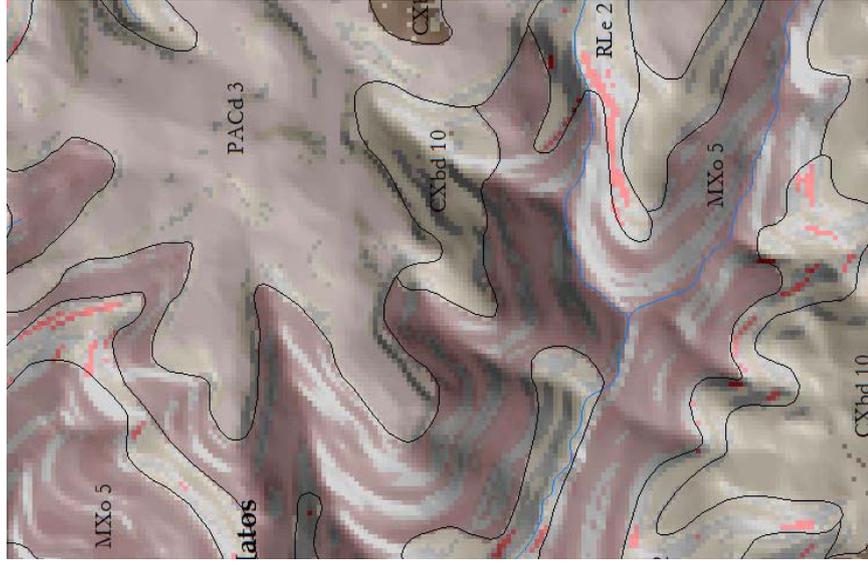
borda de algumas UM originais ainda permanece perceptível em certas áreas, o que não se verifica no mapa obtido com o procedimento de desagregação livre (Figura 12c). De fato, a desagregação livre tende a ser mais realista porque cada *pixel* pode ser alocado a qualquer uma das classes, mesmo que ela não tenha sido descrita para determinada UM, possibilitando assim mapear inclusões. O pressuposto é que condições ambientais semelhantes tendem a resultar em solos semelhantes, assim como no delineamento manual, com a vantagem da consistência na aplicação das regras (Roecker et al., 2010). Bui & Moran (2001) consideraram que a harmonização bem sucedida entre bordas de UM adjacentes confere alguma confiança adicional ao resultado da desagregação.

Entretanto, frequentemente o que define se um mapa é apropriado é sua validade para uso prático, tais como na estimativa de propriedades dos solos ou em avaliações de aptidão dos solos para diferentes usos (Rossiter, 2004). Em mapas convencionais os modelos ou critérios de avaliação geralmente são aplicados a toda a UM, tomando como base as propriedades da classe de solo dominante ou um valor ponderado pelas frações das diferentes classes. Mapas de classes de solo desagregadas podem ser mais efetivos, pois a localização provável das classes individuais já está definida e os modelos ou critérios podem ser aplicados a cada uma independentemente (Nauman et al. 2012; Kerry et al., 2012; Li et al, 2012; Odgers et al., 2014b).

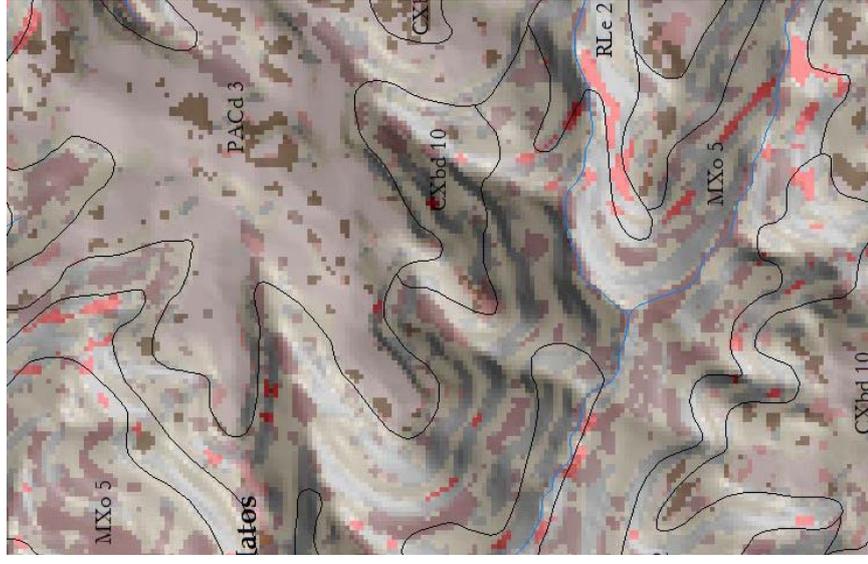
Em comparação com uma avaliação de aptidão para a viticultura realizada previamente sobre o mapa convencional de solos (Flores et al., 2007d; Figura 13a), os mapas desagregados (Figura 13b e Figura 13c) destacam áreas preferenciais e recomendadas dentro de UM que haviam sido enquadradas como não recomendadas com base na classe dominante. A proporção considerável de áreas aptas ao uso em questão em UM anteriormente consideradas inaptas mostra o quanto o desconhecimento da localização das classes pode prejudicar o resultado, pois uma série de áreas viáveis seria desconsiderada. O contrário é válido para UM anteriormente consideradas preferenciais ou recomendadas. Adicionalmente, ambos os mapas de aptidão desagregados são visualmente mais coerentes com o uso atual na área de estudo (Figura 13d), onde predomina a viticultura, que o mapa de aptidão derivado das UM originais.



(a)



(b)



(c)

Figura 12. Detalhe de uma porção da área de estudo: (a) UM combinadas no mapa convencional original; (b) classes desagregadas restringindo a predição pelo limite de cada UM; (c) classes desagregadas sem restringir a predição pelo limite de cada UM (Os limites das UM foram mantidos nos dois últimos para ilustrar a distribuição espacial das classes em relação aos polígonos do mapa original).

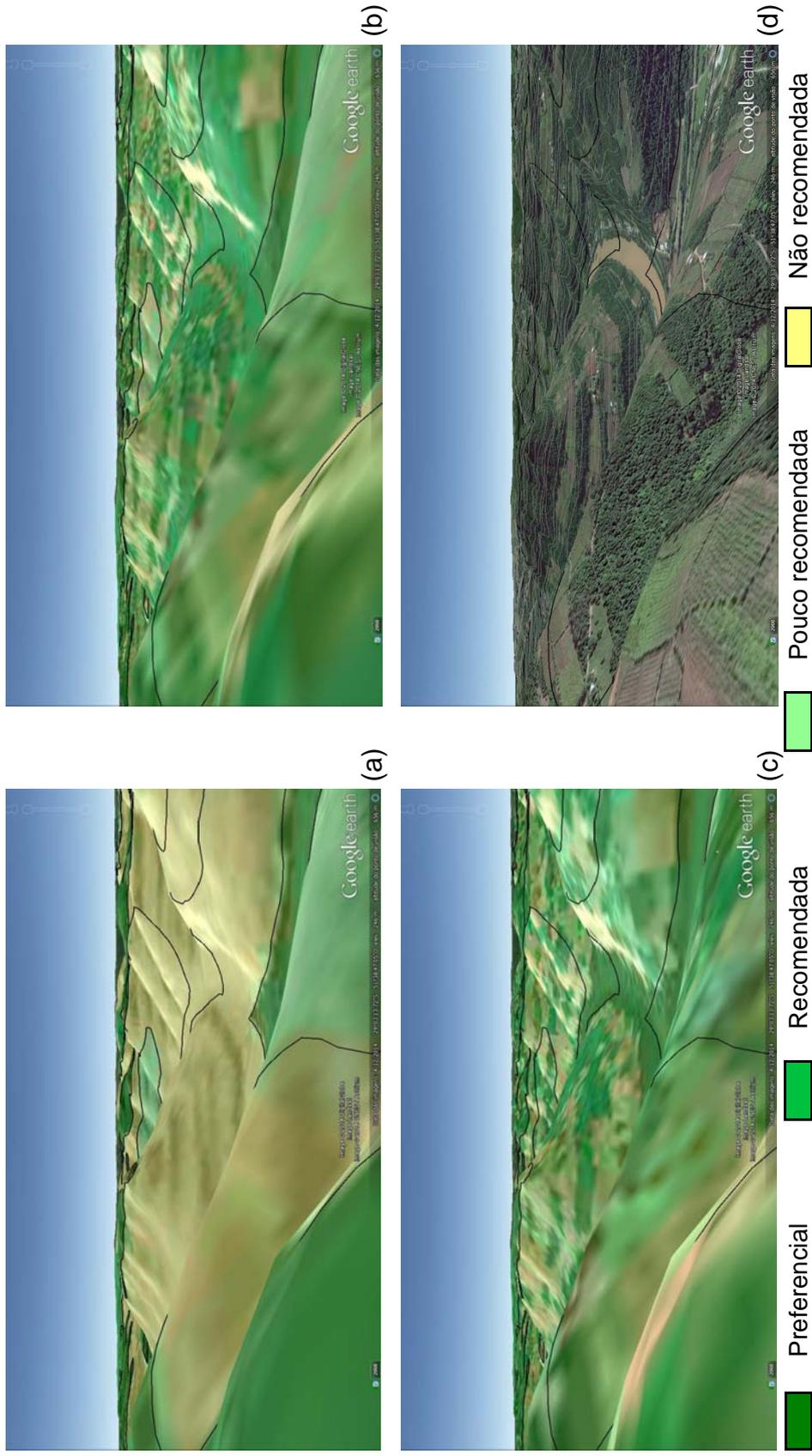


Figura 13. Visão 3D no Google Earth com o resultado da avaliação da aptidão dos solos à viticultura (Flores et al., 2007d): (a) nas UM do mapa convencional, com base na classe de solo dominante; (b) nas classes desagregadas restringindo pelos limites das UM; (c) nas classes desagregadas sem restringir pelos limites das UM; (d) imagem ilustrando o uso do solo, mais coerente com as classes desagregadas do que os limites das UM (linhas pretas).

Comparados entre si, os mapas desagregados de aptidão à viticultura (Figura 13b e Figura 13c) são mais semelhantes do que o observado na comparação da distribuição espacial das classes de solo individualizadas com os dois procedimentos utilizados (Figura 12b e Figura 12c). Isso ocorreu porque classes taxonômicas diferentes, mas com certas propriedades equivalentes, foram enquadradas na mesma classe de aptidão, de acordo com os critérios de avaliação. Tal resultado revela que algumas medidas usadas para avaliação de acurácia podem subestimar a validade de mapas desagregados para certas aplicações. Por exemplo, a predição de cambissolo no lugar de neossolo em uma UM alterou as frações e áreas em relação ao relatório de levantamento e reduziu a EG devido a discordâncias com alguns *pedons* independentes, mas é uma discordância que representa baixo risco para a tomada de decisão sobre o uso do solo (Rossiter, 2004).

4.3.4. Considerações finais

Os resultados obtidos com a metodologia de desagregação proposta mostram que é possível individualizar as classes de solo de UM de mapas convencionais com base em conhecimento especialista e técnicas de MDS. Com relação ao mapa convencional utilizado, a escala cartográfica efetiva consideravelmente inferior à nominal e a riqueza temática da legenda forneceram os primeiros indicativos nesse sentido, e podem ser um critério para avaliar a possibilidade de desagregação de outros mapas (Sarmiento et al., 2014). Os mapas resultantes da desagregação apresentaram uma melhora substancial no detalhe espacial em relação ao mapa convencional original, para o máximo grau que os dados disponíveis permitem, sem perda de acurácia ou necessidade de novas amostragens em campo. A individualização de áreas típicas com base em conhecimento, embora trabalhosa, possibilitou estabelecer critérios claros de mapeamento e provou ser útil no desenvolvimento de estratégias consistentes para delinear não apenas solos, mas diversos recursos naturais (Fisk et al., 2010).

As informações descritivas vagas ou com lacunas no relatório de levantamento de solos constituíram uma dificuldade, mas essa é uma situação comum em muitos países, (Bui & Moran, 2001; Odgers et al., 2014a). O procedimento aqui utilizado buscou complementá-las mediante observações em campo, mas pedólogos experientes poderiam estimar muitas das

informações necessárias diretamente, como demonstrou a colaboração de um dos autores do levantamento de solos original. A disponibilidade de *pedons* independentes para a validação dos mapas desagregados geralmente também é muito restrita, mas os resultados mostraram que as frações e áreas das classes e uma análise visual podem suprir medidas satisfatórias sobre sua qualidade. Quando possível, pode-se considerar a aquisição de pontos em campo com a identificação das classes de solo até onde é possível a partir da descrição morfológica.

A disponibilidade limitada de dados é uma realidade na qual o uso de conhecimento especialista apresenta vantagens em relação a abordagens estatísticas (Smith et al., 2012; Ashtekar & Owens, 2014; Odgers et al., 2014a). No caso da desagregação de mapas convencionais, essa abordagem também representa uma oportunidade de incorporar a experiência de muitos pedólogos em atividade para gerar informações de solos mais detalhadas a baixo custo. Muitos desses pedólogos não estão familiarizados com técnicas de MDS, mas podem colaborar para extrair o máximo dessa valiosa ferramenta a serviço da ciência do solo. Além disso, a introdução de conhecimento especialista possibilita capturar o que nenhum modelo matemático consegue substituir, a capacidade da mente humana de identificar sutilezas e formular hipóteses (Ashtekar & Owens, 2014).

Além de representar com maior detalhe a distribuição espacial de classes de solo, a desagregação de mapas convencionais pode servir a muitos outros propósitos. Mapas desagregados podem servir de ponto de partida para atualizar mapas existentes e planejar novos levantamentos em campo. Eles também podem ser usados em aplicações desenvolvidas diretamente sobre as classes desagregadas ou como dados de entrada para o desenvolvimento de modelos de MDS para prever a distribuição espacial de propriedades dos solos. Finalmente, é importante ressaltar que, diferente de um mapa convencional, mapas desagregados não constituem um produto acabado, eles são versões de diferentes resultados que podem ser obtidos a partir dos mesmos dados originais. Novas versões podem ser criadas a qualquer momento, dependendo do propósito e das técnicas de MDS, ou das covariáveis explanatórias utilizadas.

4.4. Conclusões

- i. O uso de técnicas de MDS e de conhecimento especialista para a desagregação de polígonos em mapas convencionais de solos no Brasil é viável e promissora, possibilitando resultados com maior detalhe espacial e acurácia igual ou superior à do mapa original;
- ii. A identificação de áreas de ocorrência típicas de cada classe de solo na área de interesse é crítica para o resultado, e pressupõe conhecimento local sobre as relações solo-paisagem;
- iii. O detalhamento vago de tais relações usualmente verificada nos relatórios de levantamento torna altamente recomendável a realização de viagens de reconhecimento para complementar ou consolidar o conhecimento sobre as mesmas;
- iv. Em função das lacunas frequentes nas informações descritivas nos relatórios de levantamento, métodos de desagregação automatizados ou semi-automatizados para desagregação de unidades de mapeamento tendem a ser de aplicação mais difícil.

5. CAPÍTULO IV – PREDIÇÃO DE CLASSES DE SOLOS COM ALTA RESOLUÇÃO ESPACIAL EM ÁREAS NÃO AMOSTRADAS A PARTIR DE UM MAPA DETALHADO EXISTENTE

5.1. Introdução

A demanda por informações detalhadas sobre os solos tem crescido consideravelmente nas últimas décadas, e tende a se intensificar à medida que aumenta a pressão sobre os recursos naturais e se acentua a necessidade de preservação do ambiente para a sustentação da vida (Håring et al., 2012; Miller, 2012; Omuto et al., 2013). Os levantamentos convencionais de solos, executados de forma mais ou menos sistemática há mais de um século, ainda são a fonte mais importante de informações sobre as características e a distribuição desse corpo natural (Carré et al., 2007). No entanto, a maioria dos levantamentos existentes foi executada com baixo nível de detalhamento, resultando em mapas de solos com escalas pequenas (McBratney et al., 2003; Sanchez et al., 2009). Hartemink (2008), por exemplo, estimou que mapas de solos em escalas maiores que 1:100.000 cobrem somente 10% a 30% dos países da América do Sul, sudeste da Ásia e da África.

O mapeamento preditivo de solos, hoje genericamente denominado de mapeamento digital de solos (MDS), surgiu como alternativa para tentar suprir essas novas demandas, visando preencher lacunas, ampliar a cobertura do mapeamento e aumentar a resolução das informações sobre a distribuição espacial dos tipos ou de certas propriedades dos solos (McBratney et al., 2003; Scull et al., 2005). Segundo Lagacherie & McBratney (2007), o MDS 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 sobre os solos e de covariáveis ambientais relevantes.

No contexto de MDS, os levantamentos convencionais existentes adquirem importância como fonte de informações para subsidiar a calibração dos modelos de predição de classes ou de propriedades dos solos (Scull et al., 2005; Omuto et al., 2013). Diferentes estratégias têm sido propostas nesse sentido, levando em conta as características e limitações do material existente em cada local. Em áreas que possuem cobertura completa com mapas e perfis de solos, uma das estratégias recomendadas é a extrapolação (Minasny & McBratney, 2010). Ela adota um raciocínio similar ao usado em um levantamento convencional, onde observações feitas em alguns pontos no terreno são depois usadas para inferir e delinear a ocorrência dos solos na paisagem em toda a área de interesse (Bui, 2004).

A extrapolação propõe o uso de dados de solos de uma área representativa das relações solo-paisagem para criar regras de mapeamento com base em covariáveis ambientais disponíveis, e então prever classes ou propriedades de solos em áreas não mapeadas com características ambientais semelhantes (Fawcett et al., 1993; Lagacherie et al., 1995; Bui & Moran, 2001; Grinand et al., 2008; Lemercier et al., 2012; Cambule et al., 2013; Höfig et al., 2014). O processo envolve a recuperação do modelo mental de solo-paisagem implícito nos mapas convencionais, usando covariáveis que representam os fatores de formação dos solos (Bui, 2004). As árvores de decisão são eficazes para esse propósito devido à sua capacidade de processar dados contínuos e discretos, independência de suposições sobre a sua distribuição, facilidade de caracterizar relações não lineares e obtenção de regras de decisão interpretáveis (Bui & Moran, 2001; Qi & Zhu, 2003; Subburayalu & Slater, 2013; Cavazzi et al., 2013; Lorenzetti et al., 2015).

Algumas das aplicações atuais exigem informações detalhadas sobre os solos, entre elas a execução de obras civis, a deposição de resíduos e a produção agrícola de alto valor agregado, com destaque para a viticultura (Bragato, 2004). No entanto, levantamentos detalhados de solos só existem para porções muito restritas na maioria dos países por causa dos altos custos envolvidos, e frequentemente estão dispersos em pequenas áreas. Por exemplo, Lagacherie et al. (1995) estimaram que apenas 10% da França estavam cobertos por mapas detalhados de solos, apesar da sua longa tradição vitivinícola e da reconhecida importância do solo na qualidade dos

produtos dessa atividade. Em tais áreas, a extrapolação torna-se uma alternativa extremamente importante, pois representa a oportunidade de ampliar a cobertura do mapeamento de solos com alta resolução espacial a baixo custo.

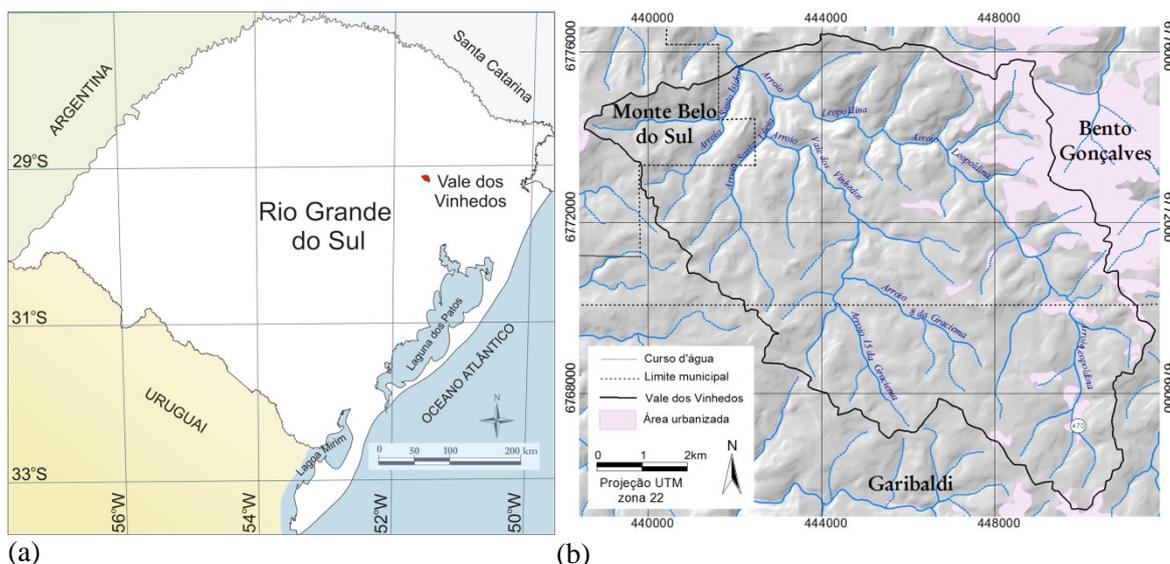
O objetivo principal deste estudo foi avaliar a possibilidade de efetuar a predição de classes de solos em áreas não visitadas, ou seja, a extrapolação, a partir de um levantamento convencional detalhado existente na vizinhança e de covariáveis explanatórias com alta resolução espacial. Os objetivos secundários foram avaliar a contribuição de diferentes covariáveis e o impacto da amostragem e da resolução espacial sobre o desempenho da predição. A hipótese é que o modelo solo-paisagem implícito no mapa existente pode ser recuperado e reaplicado para predizer classes de solo em áreas vizinhas na mesma paisagem, com base em um conjunto de covariáveis ambientais que possibilite descrever os principais fatores de formação dos solos na região.

5.2. Material e métodos

5.2.1. Área de estudo

A área de estudo compreende os limites da Indicação de Procedência (IP) Vale dos Vinhedos, situada na zona de produção vitivinícola Serra Gaúcha, nordeste do estado do Rio Grande do Sul, aproximadamente entre as latitudes 29°14'26,51"S e 29°8'23,67"S e as longitudes 51°37'57,31"W e 51°29'50,83"W (Figura 14). Sua superfície totaliza 8.121,59 hectares e divide-se entre os municípios de Bento Gonçalves (aproximadamente 61% da área), Garibaldi (cerca de 34%) e Monte Belo do Sul (cerca de 5%).

De acordo com a classificação de Köppen, o clima no Vale dos Vinhedos enquadra-se no tipo Cfb, subtropical úmido com verões amenos, o qual é característico das áreas de maior altitude no Rio Grande do Sul. A precipitação pluvial anual totaliza 1.736 mm (Embrapa, 2008), com o menor acumulado em maio e o maior em setembro, mas sem caracterizar um período seco e outro chuvoso. A temperatura, por outro lado, apresenta variação sazonal com verão e inverno bem definidos, sendo janeiro o mês mais quente e junho o mais frio.



(a) (b)
 Figura 14. (a) Localização do Vale dos Vinhedos no estado do Rio Grande do Sul e; (b) limites do Vale dos Vinhedos.

Quanto à geologia, a maior parte do Vale dos Vinhedos situa-se sobre a sequência superior da Formação Serra Geral (fácies Caxias), constituída de rochas efusivas ácidas. Uma fração menor da área, a noroeste, situa-se sobre a sequência inferior (fácies Gramado), de composição predominantemente básica (Flores et al., 2012). O Vale dos Vinhedos insere-se na região geomorfológica Planalto das Araucárias e na região fisiográfica da Encosta Superior do Nordeste (IBGE, 1986). Seu relevo é heterogêneo, com altitudes entre 225m e 720m e grandes variações de declividade e de exposição solar. A topografia é ondulada no topo e apresenta bordas escarpadas e recortadas, com drenagem fortemente encaixada. Os vales mais profundos estão situados a noroeste, junto ao curso inferior dos arroios Leopoldina, Vale dos Vinhedos, Santa Lúcia e Santo Isidoro, em cujas encostas encontram-se as áreas mais íngremes.

A distribuição dos solos é fortemente associada ao relevo e, assim como este, é altamente heterogênea, observando-se uma predominância relativa de solos rasos e pedregosos (Flores et al., 2012). A vegetação original, predominantemente de floresta subtropical decídua, foi altamente modificada a partir de 1874, com o início da colonização italiana (IBGE, 1986). Atualmente a principal atividade agropecuária é a vitivinicultura, desenvolvida em pequenas propriedades com pouca mecanização e predomínio da mão-de-obra familiar (Protas et al., 2006).

A escolha do Vale dos Vinhedos como área de estudo deve-se à disponibilidade de um levantamento detalhado de solos e à demanda por informações mais detalhadas sobre solos em duas áreas vizinhas: Monte Belo do Sul e Pinto Bandeira. O motivo é a busca dessas áreas pelo registro de Indicações Geográficas (IG) para vinhos, seguindo o caminho trilhado pelo Vale dos Vinhedos a partir de 1999. Ambas já adquiriram bases cartográficas apropriadas, mas a execução de novos levantamentos de solos enfrenta dificuldades logísticas e de recursos. Nesse contexto, o uso do levantamento detalhado existente como referência para predizer classes ou propriedades de solos em tais áreas representa uma alternativa para suprir as informações de solos demandadas.

5.2.2. Material utilizado

O material utilizado neste estudo compreendeu um levantamento detalhado de solos, incluindo o respectivo relatório de levantamento, um mapa convencional com escala de apresentação 1:10.000 e um conjunto de 163 perfis e 74 pontos de observação de solos, com coordenadas registradas via GPS (Flores et al., 2012). Também foi utilizada uma base cartográfica digital compatível com a escala 1:5.000, contendo camadas vetoriais da rede hidrográfica, sistema viário e curvas de nível (equidistância vertical de 5 metros), além de um mosaico de fotografias aéreas ortorretificadas com resolução espacial de 60 cm. Todos os dados geoespaciais foram estruturados em coordenadas UTM, fuso 22, referenciadas ao datum SIRGAS2000. Para o processamento e análise foram utilizados os programas ArcGIS (ESRI, 2013) e Idrisi (Eastman, 2012), além de equações e rotinas disponibilizadas gratuitamente (Finnen & Menza, 2013).

5.2.3. Metodologia

A metodologia utilizada usa uma abordagem de mineração de dados para recuperar o conhecimento implícito em um mapa convencional de solos existente e transcrevê-lo para o espaço de covariáveis explanatórias, para então aplicá-lo na predição de classes de solos em áreas adjacentes não mapeadas a partir dessas covariáveis (Lagacherie et al., 1995; Grinand et al. 2008; Lemercier et al., 2012; Cambule et al., 2013). Ela também adota um princípio de repetição para realizar diversas predições das mesmas classes,

visando aumentar a confiabilidade dos resultados e, adicionalmente, quantificá-la de forma empírica (Bui & Moran, 2001; Wei et al., 2010; Nauman & Thompson, 2014). A Figura 15 apresenta uma visão geral da metodologia utilizada, cujos passos são descritos em maior detalhe a seguir.

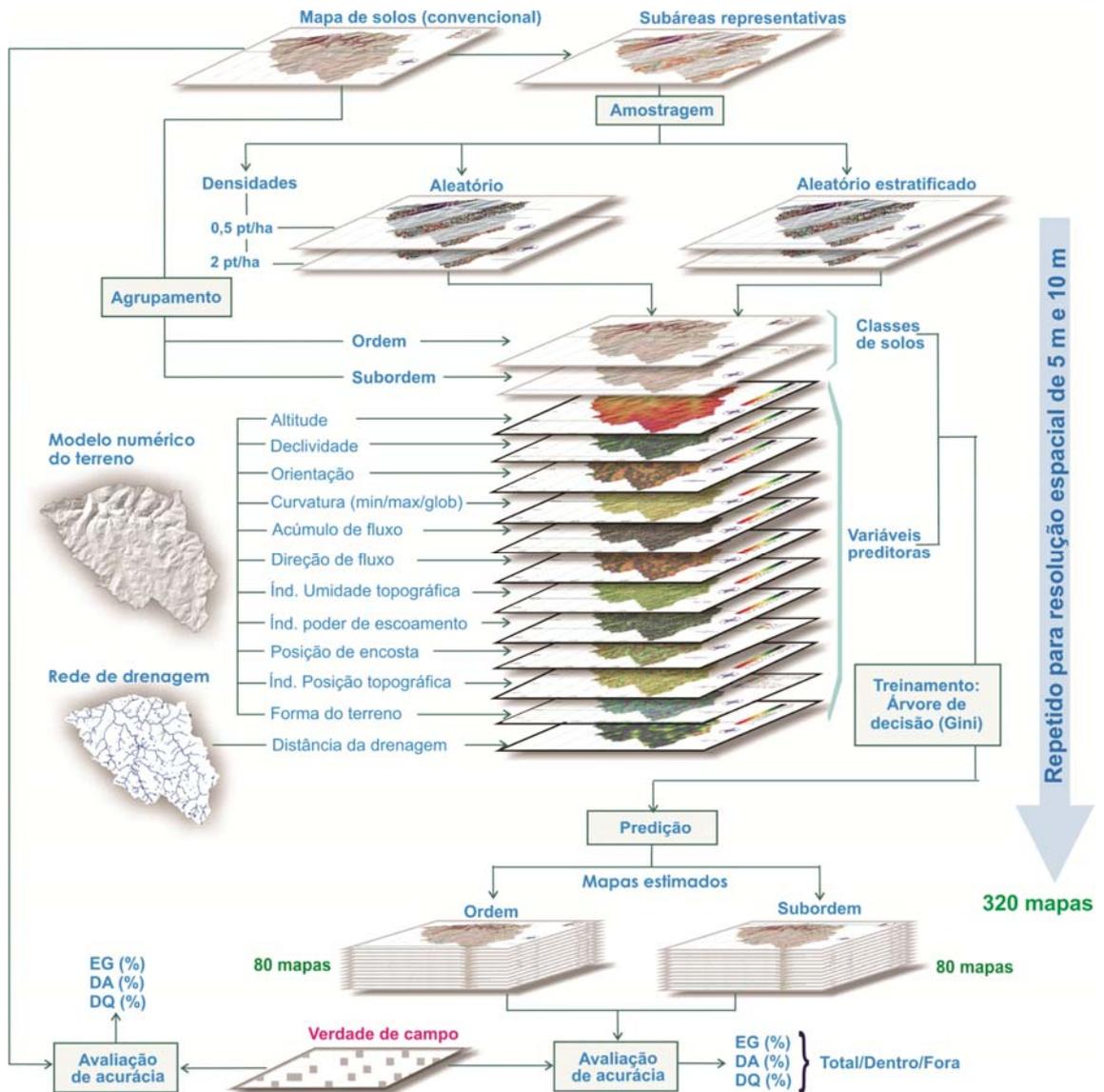


Figura 15. Fluxograma da metodologia utilizada na predição de classes de solos em áreas não amostradas no Vale dos Vinhedos.

Com vistas a simular a predição em áreas não mapeadas a partir de áreas de referência, a área de estudo foi dividida em três subáreas representativas das diferentes condições de relevo e relações solo-paisagem na região (Figura 16): (1) área de maior altitude e com relevo mais suave, onde predominam solos mais desenvolvidos; (2) área de altitude intermediária, com condições de relevo mais heterogêneas e maior diversidade de tipos de solos;

(3) área de vale em menor altitude, com relevo acidentado, onde predominam solos rasos e pedregosos. Essas subáreas totalizam 43% da superfície total, aproximadamente a mesma proporção da superfície total do Vale dos Vinhedos em relação à das duas áreas adjacentes não mapeadas anteriormente citadas.

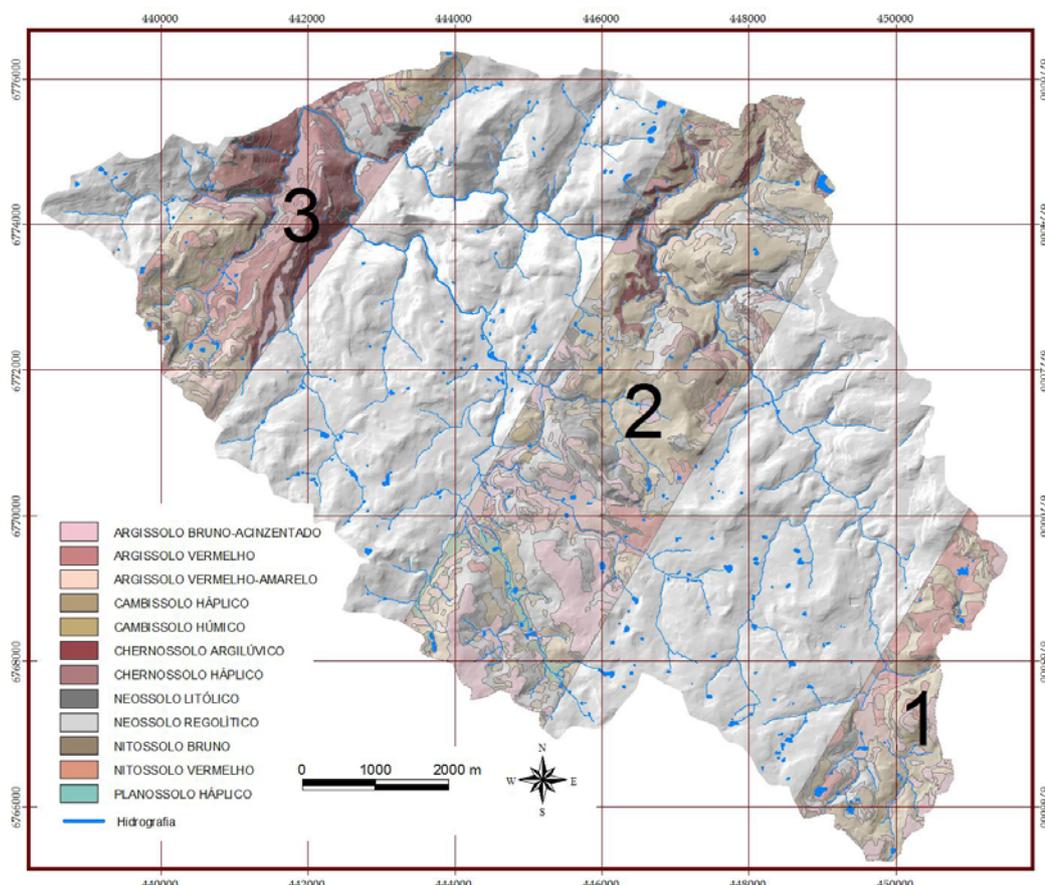


Figura 16. Subáreas representativas do Vale dos Vinhedos utilizadas para obtenção de amostras de treinamento (1 = área de maior altitude e relevo mais suave; 2 = área de altitude intermediária e relevo heterogêneo; 3 = área de menor altitude e relevo acidentado).

As subáreas representativas selecionadas foram usadas como perímetro de amostragem, considerando-se a distribuição espacial das classes taxonômicas de solo em nível de ordem (6 classes) e de subordem (12 classes). Para ambos os níveis de classificação, foram gerados conjuntos de pontos amostrais com as densidades de 0,5 e 2 pontos ha^{-1} , resultando respectivamente em 1.740 pontos amostrais e 7.040 pontos. As duas densidades estão dentro das recomendações para levantamentos detalhados de solos (IBGE, 2007) e também constituíram extremos de um intervalo com resultados igualmente satisfatórios na predição de classes de solo em uma porção menor da área de estudo (Giasson et al., 2011; Sarmento et al., 2012). A opção pelas duas buscou verificar se a variabilidade nas relações solo-

paisagem decorrente da expansão da área em relação a esses estudos prévios é suficientemente caracterizada com a menor densidade ou se eventualmente demanda mais pontos amostrais.

Para cada nível de classificação taxonômica e cada densidade de amostragem, os pontos foram gerados adotando-se dois esquemas de distribuição espacial: totalmente aleatório e aleatório estratificado (Finnen & Menza, 2013). Com vistas a reduzir a introdução de incerteza na caracterização das classes, durante a geração dos pontos de amostragem excluiu-se um *buffer* de 10 m ao longo das bordas dos polígonos do mapa convencional de solos, distância que corresponde a 1 mm na sua escala de apresentação e equivale ao erro gráfico (Grinand et al., 2008). Ao final, considerando-se os dois níveis de classificação taxonômica, as duas densidades de amostragem e os dois esquemas de distribuição espacial, foram gerados oito conjuntos de pontos amostrais.

As covariáveis para caracterizar as classes de solos foram definidas a partir do material cartográfico básico disponível no Vale dos Vinhedos e nas duas áreas adjacentes não mapeadas, no qual constam basicamente relevo e hidrografia. Ao todo, foram selecionadas 14 covariáveis, sete das quais usadas anteriormente para descrever a variação espacial de classes de solos em uma porção da área de estudo por Giasson et al., 2011 e Sarmiento et al., 2012. As demais foram acrescentadas tomando-se como base o conhecimento local sobre as relações solo-paisagem e os fatores de formação dos solos, priorizando-se covariáveis relacionadas à erosão e à deposição de sedimentos, ao regime de umidade do solo, ao teor de matéria orgânica e à espessura do horizonte superficial. A maioria das covariáveis selecionadas corresponde a medidas morfométricas do relevo, mas elas também descrevem direta ou indiretamente outros fatores de formação dos solos (Tabela 7).

Para o cálculo das covariáveis de relevo foram elaborados dois modelos numéricos do terreno (MNT), mediante uma interpolação a partir das curvas de nível constantes na base cartográfica. O método de interpolação empregado foi uma rede triangular irregular (Triangulated Irregular Network - TIN), com uma função parabólica para ajustar as quebras de relevo e estimar a altitude do fundo dos vales e do topo de elevações. A opção pela interpolação TIN levou em consideração suas vantagens na representação de relevos

complexos como o da área de estudo (Medeiros & Câmara, 2001). Redes triangulares podem incorporar facilmente descontinuidades e são estruturas eficientes porque a densidade dos triângulos pode variar para se ajustar a variações locais no terreno (Moore et al., 1991; Wilson & Gallant, 2000). Uma vez criada a estrutura TIN, as altitudes foram convertidas para o modelo *raster* (matricial) com as resoluções espaciais de 5 m e de 10 m, a partir das quais se calculou as covariáveis.

Tabela 7. Covariáveis explanatórias e fatores de formação dos solos representados pelas mesmas no Vale dos Vinhedos (C: clima, R: relevo, P: material de origem, T: tempo).

Covariável preditora	Fator de formação
Altitude	R, C, P, T
Declividade	R
Orientação da encosta	R, C
Curvatura mínima	R
Curvatura máxima	R
Curvatura global	R
Distância da rede de drenagem	R, P
Direção de fluxo	R
Acúmulo de fluxo	R
Índice de poder de escoamento (<i>Stream Power Index</i> - SPI)	R
Índice de umidade topográfica (<i>Topographic Wetness Index</i> - TWI) ¹	R
Índice de posição topográfica (<i>Topographic Position Index</i> - TPI) ¹	R
Posição na encosta (<i>Slope position</i>)	R, C
Classes de forma do terreno (<i>Landform classes</i>)	R

¹ Computado com vizinhança de 100 m e de 145m (máxima extensão possível com o MNT)

As covariáveis explanatórias foram posteriormente usadas para caracterizar as classes de solo, tomando-se os pontos amostrais como áreas de treinamento, e então predizer a ocorrência das classes em todos os *pixels* da área de estudo. Nesta etapa utilizou-se um algoritmo de árvore de decisão que emprega o índice de impureza de Gini para avaliar a qualidade das subdivisões durante a construção da árvore. A impureza de Gini é uma medida da heterogeneidade interna do conjunto de *pixels* pertencentes à classe formada em uma determinada subdivisão do conjunto de dados. Durante o processo de construção da árvore, as subdivisões são ajustadas recursivamente com vistas a minimizá-la e a criar folhas com somente uma classe, ou com um índice de impureza próximo de zero (Nauman & Thompson, 2014).

O processo de treinamento e predição foi repetido várias vezes, variando-se o número e a combinação de covariáveis explanatórias utilizadas. O conjunto inicial contemplou as sete covariáveis utilizadas por Giasson et al. (2011) e Sarmiento et al. (2012), sendo as demais acrescentadas sucessivamente. A cada acréscimo, as covariáveis foram alternadas a fim de se identificar a melhor combinação para o número de covariáveis em questão, usando como critério o menor erro de treinamento. Dessa forma, foi selecionada a melhor combinação de 8 covariáveis, a melhor de nove, e assim por diante. Cada combinação foi usada com os pontos amostrais gerados previamente com dois esquemas de distribuição espacial, duas densidades de amostragem e dois níveis de classificação dos solos, para as duas resoluções espaciais, resultando em várias árvores de decisão. Para controlar o grau de ramificação de cada árvore utilizou-se uma poda automática limitada por área, optando-se por suprimir nós terminais (folhas) com superfície inferior a 1% da área total (Bui & Moran, 2001). Os *pixels* correspondentes às folhas que não atingiram o limiar de área mínima foram recursivamente alocados a outras classes de modo a minimizar a heterogeneidade interna das mesmas, até a estabilidade dos erros de treinamento (Lippitt et al., 2008). As diversas árvores de decisão obtidas foram então empregadas para predizer as classes de solo em toda a área de estudo, cada árvore resultando em um mapa predito.

A avaliação de acurácia dos mapas resultantes da predição foi efetuada usando-se como referência 237 *pedons* do levantamento convencional de solos, sendo 163 perfis e 74 pontos de observação, todos com coordenadas coletadas via GPS (Figura 17). Adicionou-se um *buffer* de 1 *pixel* em torno de cada ponto e atribuiu-se a todas as células a classe do ponto central, aumentando para 2.124 o número de *pixels* de referência. Assumiu-se que locais muito próximos tendem a pertencer à mesma classe de solo ou a ter propriedades similares devido à autocorrelação espacial (Smith et al., 2012). O uso de um *buffer* também visou detectar a transição gradual entre classes que ocorre na realidade e de compensar eventuais incertezas remanescentes do delineamento dos polígonos do mapa convencional (Smith et al., 2012; Nauman & Thompson, 2014).

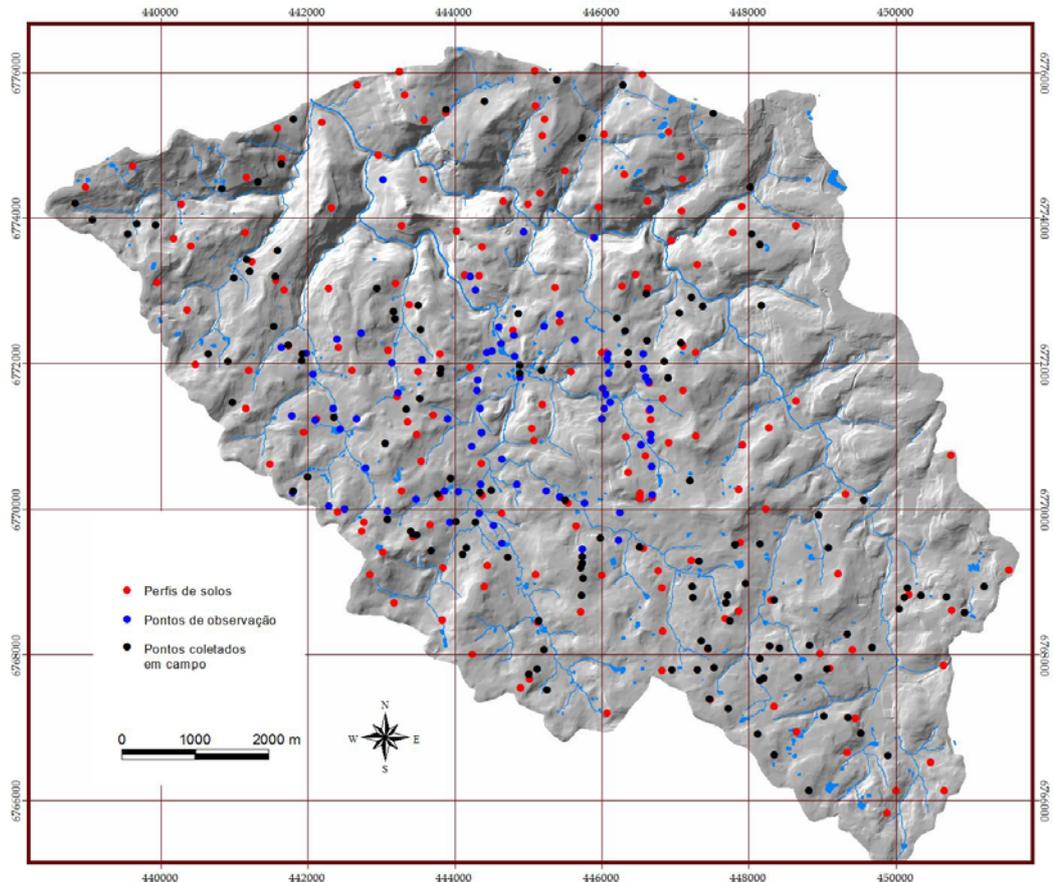


Figura 17. Localização de 237 *pedons* existentes (perfis e pontos de observação), usados para a validação dos mapas preditos, e de 130 *pedons* independentes coletados em campo para a validação do mapa convencional.

Para cada mapa predito, computou-se uma matriz de erros (Congalton, 1991) e quantificou-se a exatidão global (EG, percentagem de acertos), e as discordâncias de quantidade (DQ) e de alocação (DA) das classes preditas (Pontius & Millones, 2011). Cada um desses indicadores de acurácia foi computado separadamente para três áreas de abrangência diferentes, quais sejam: 1) área de treinamento, constituída pelas três subáreas representativas; 2) porção fora da área de treinamento e; 3) extensão total da área de estudo. Adicionalmente, para o melhor mapa obtido em cada nível de classificação taxonômica e em cada resolução espacial, calculou-se a frequência relativa da classe predita em cada *pixel*. O cálculo foi efetuado contando-se o número de vezes em que a classe foi repetida no mesmo *pixel* ao longo de todo o conjunto de mapas e dividindo-se essa contagem pelo número de mapas. O resultado varia de 0 a 1 e pode ser equiparado à probabilidade de ocorrência da classe, servindo como uma aproximação empírica para avaliar a confiança na predição (Bui & Moran, 2001; Wei et al., 2010; Nauman & Thompson, 2014).

Para subsidiar a análise e interpretação da avaliação de acurácia da predição, foi também realizada uma avaliação do mapa convencional de solos usado no treinamento do algoritmo de classificação. Utilizou-se 130 pontos independentes coletados em campo, cuja localização foi definida por sorteio, optando-se por uma distribuição espacial totalmente aleatória (Finnen & Menza, 2013) (Figura 17). Com o auxílio de receptores GPS, cada ponto foi visitado *in loco*, onde foi realizada a identificação da classe de solo em nível de ordem e de subordem a partir das características morfológicas dos perfis. Os pontos com as respectivas classes foram então empregados para calcular para o mapa convencional os mesmos indicadores de acurácia descritos para os mapas resultantes da predição.

5.3. Resultados e discussão

Das 14 covariáveis explanatórias inicialmente relacionadas na Tabela 7, a última foi descartada por não modificar a EG dos mapas preditos quando incluída. As covariáveis restantes possibilitaram a obtenção de 20 diferentes combinações, com número de covariáveis variando de sete a 12. As combinações entre covariáveis preditoras, densidades de amostragem, esquemas de distribuição espacial dos pontos amostrais, número de covariáveis explanatórias e resolução espacial, possibilitaram a predição de 320 mapas de classes de solos. Desse total, 160 têm resolução espacial de 5 m e outros 160 possuem resolução de 10 m. Em cada caso, 80 correspondem a mapas com classes preditas em nível de ordem e 80 com classes em nível de subordem.

A análise de variância da EG dos mapas preditos em nível de ordem e de subordem foi significativa a 1% de probabilidade para ambos os níveis de classificação, com F calculado consideravelmente maior que F crítico. Considerando-se a área total mapeada, os mapas de classes em nível de ordem obtidos a partir de pontos amostrais com densidade de 0,5 pontos.ha⁻¹ apresentaram EG sempre superior à dos mapas obtidos com 2 pontos.ha⁻¹ (Figura 18). A diferença em favor da menor densidade de amostragem foi significativa a 5% de probabilidade pelo teste de Tukey, e foi mais acentuada nos mapas com resolução espacial de 10 m (Figura 18c e Figura 18d). O número de covariáveis explanatórias utilizadas, por outro lado, apesar de introduzir diferenças perceptíveis, não foi estatisticamente significativa. Isso

pode ser constatado pelo fato de que a maior EG na predição de classes de solo em nível de ordem (49,61%) foi obtida com 0,5 pontos.ha⁻¹, amostragem aleatória estratificada, resolução espacial de 5 metros e 11 covariáveis explanatórias, enquanto a segunda maior EG (48,44%) foi obtida com igual amostragem e resolução espacial, porém com apenas 7 covariáveis (Figura 18b).

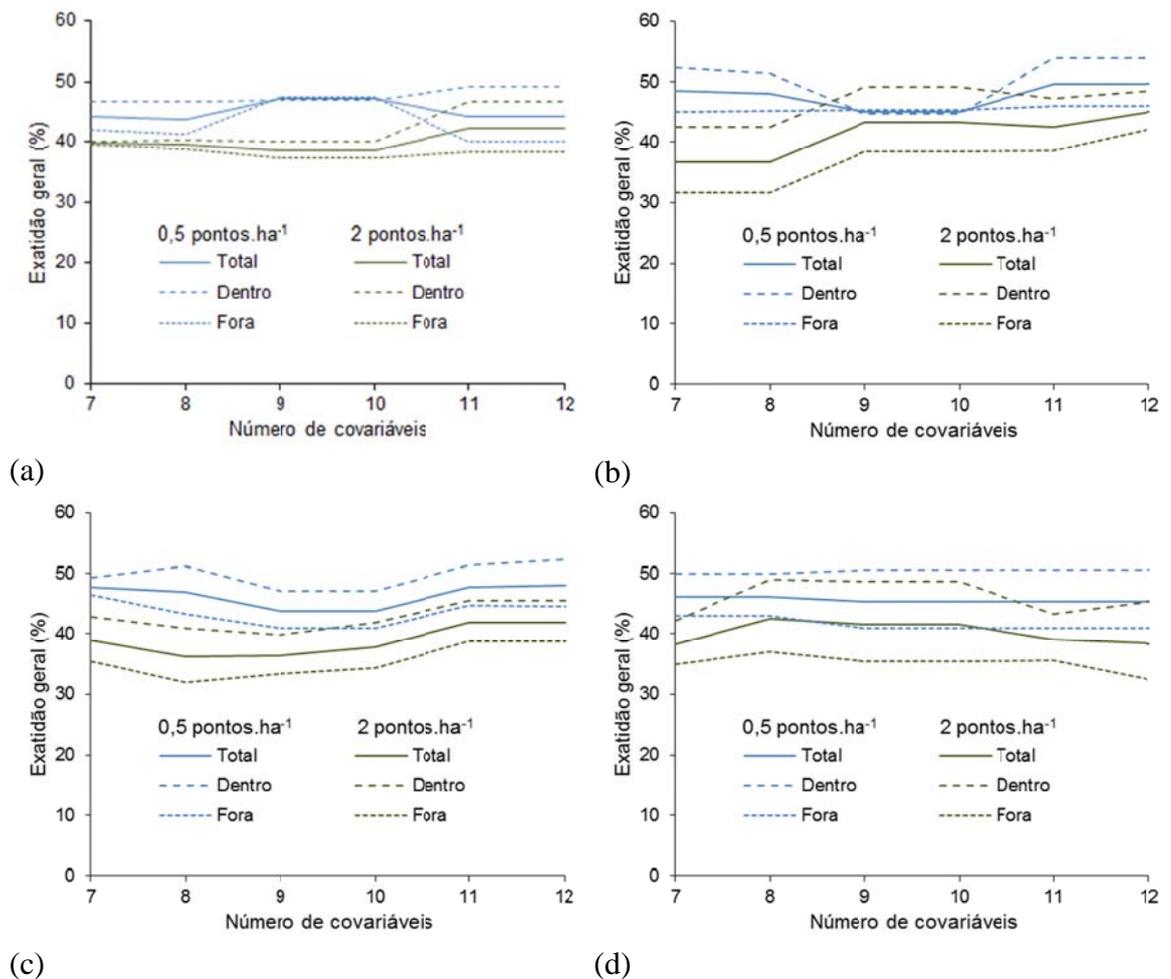


Figura 18. Exatidão geral da predição de classes de solo em nível de ordem na área total, dentro da área de treinamento e fora da área de treinamento, na resolução espacial de 5 m com amostragem (a) aleatória e (b) aleatória estratificada; e na resolução espacial de 10 m com amostragem (c) aleatória e (d) aleatória estratificada.

Em geral, os resultados obtidos para os mapas de classes previstas em nível de subordem (Figura 19) na área total mapeada reproduziram aqueles verificados para o nível de ordem, todavia com menor magnitude nos valores de EG. Em parte a redução na acurácia da predição de subordens em relação à de ordens deve-se ao incremento no número de classes, aumentando a chance de discordâncias em relação aos pontos de validação. Contudo, a redução na EG de um nível para outro foi proporcionalmente menor que a

razão entre o número de subordens (12) e o de ordens (6), o que é um aspecto positivo. Embora o número de covariáveis não tenha sido estatisticamente significativo, constatou-se que o uso de mais covariáveis tendeu a elevar a EG na predição de subordens. A maior EG (35,46%) foi obtida com 0,5 pontos.ha⁻¹, amostragem aleatória estratificada, resolução de 10 m e 11 covariáveis explanatórias, porém com prejuízo fora da área de treinamento (Figura 19d). O mapa com a segunda maior EG (34,55%) foi obtido com igual amostragem, resolução espacial de 5 m e 12 covariáveis explanatórias, neste caso com a vantagem da EG mais alta fora da área de treinamento.

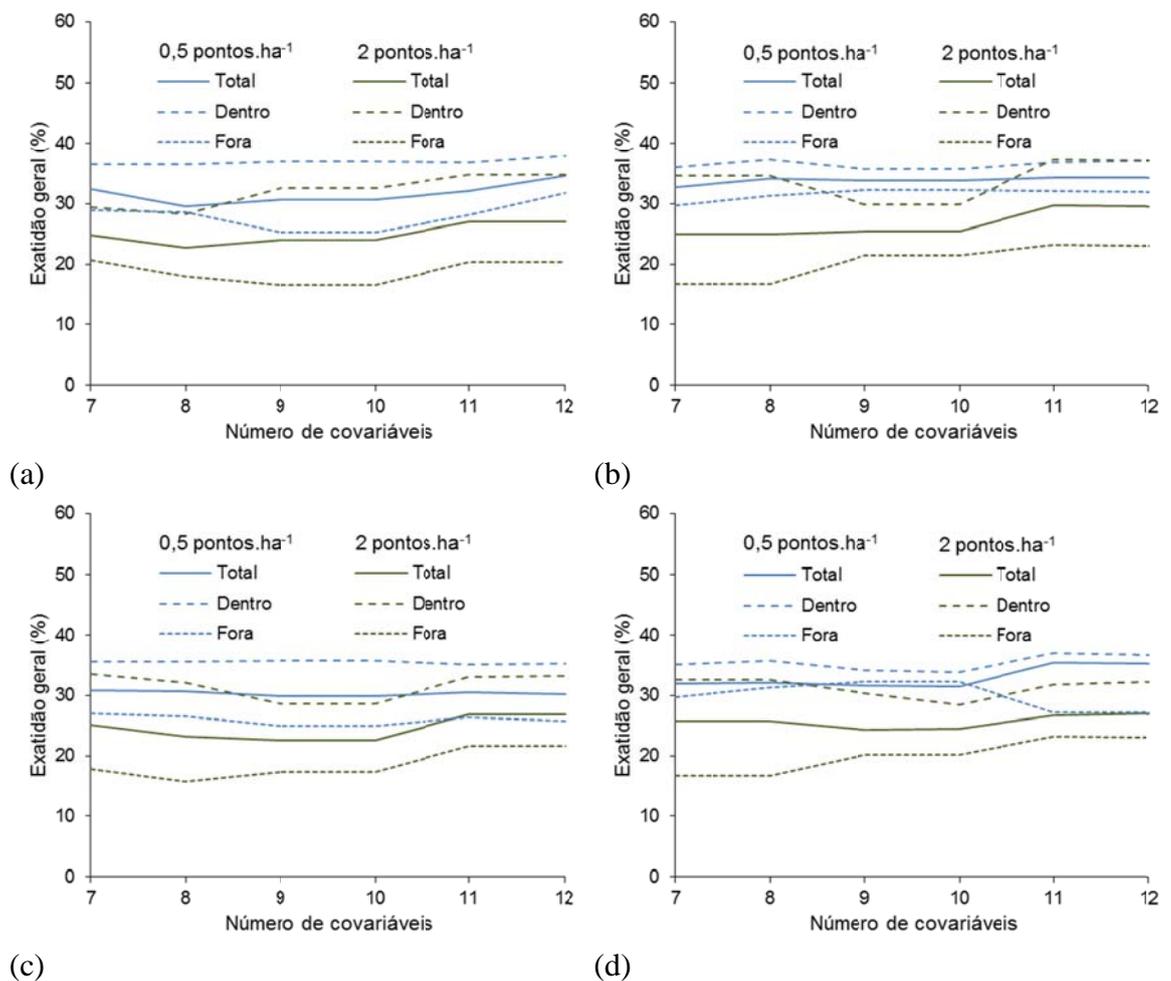


Figura 19. Exatidão geral da predição de classes de solo em nível de subordem na área total, dentro da área de treinamento e fora da área de treinamento, na resolução espacial de 5 m com amostragem (a) aleatória e (b) aleatória estratificada; e na resolução espacial de 10 m com amostragem (c) aleatória e (d) aleatória estratificada.

Os diferentes arranjos entre as duas densidades de amostragem, os dois esquemas de distribuição espacial e as duas resoluções espaciais resultaram em diferenças significativas para a EG a 5% de probabilidade, nos dois níveis de classificação taxonômica dos solos (Tabela 8). Em nível de

ordem, a maior média correspondeu aos mapas preditos com 0,5 pontos.ha⁻¹ com distribuição espacial aleatória estratificada e resolução espacial de 5 m. A mesma combinação de densidade e de esquema de amostragem também possibilitou a maior EG média na predição de subordens, porém neste nível de classificação taxonômica não houve distinção entre as resoluções espaciais de 5 m e 10 m.

Tabela 8. Comparação da EG média de mapas de classes de solo em nível de ordem e subordem, preditos a partir de covariáveis com duas resoluções espaciais e pontos amostrais gerados segundo dois esquemas de distribuição espacial e duas densidades de amostragem.

Resolução espacial	Tratamentos		Nível de classificação	
	Esquema de amostragem	Densidade (pontos.ha ⁻¹)	Ordem	Subordem
5 m	Aleatório	0,5	0,43 b	0,29 b
		2	0,38 c	0,23 d
	Aleatório estratificado	0,5	0,46 a	0,33 a
		2	0,38 c	0,26 c
10 m	Aleatório	0,5	0,45 ab	0,29 b
		2	0,38 c	0,23 d
	Aleatório estratificado	0,5	0,45 ab	0,32 a
		2	0,38 c	0,25 cd

Médias seguidas da mesma letra não diferem estatisticamente pelo teste de Tukey, no nível de 5% de significância.

A percentagem de acertos verificada para os mapas preditos nos dois níveis de classificação taxonômica (Tabela 9 e Tabela 10) encontra-se abaixo do esperado, especialmente pelo fato de que os dados de treinamento são oriundos de um levantamento detalhado de solos. Apesar disso, um aspecto positivo é que em todos os casos a maior parte das discordâncias observadas entre os mapas e os *pedons* usados para validação é de alocação (DA), restando uma parcela reduzida de discordância de quantidade (DQ). DA maior que DQ é preferível porque significa que parte dos *pixels* de algumas classes preditas não coincide exatamente com a localização dos pontos de validação, mas que a superfície das classes tende a se aproximar da real (Pontius & Millones, 2011). O que ocorreu é que a classe correta pode ter sido predita em um *pixel* situado na vizinhança do respectivo ponto de validação, em vez de exatamente na sua localização.

Tabela 9. Indicadores de acurácia do melhor mapa predito e média de todos os mapas preditos com classes taxonômicas em nível de ordem na área total mapeada, dentro da área de treinamento e fora da área de treinamento.

Indicador	Melhor predição			Média geral (160 mapas)		
	Área total	Área de treinamento	Área não amostrada	Área total	Área de treinamento	Área não amostrada
EG (%)	49,6	53,9	46,0	41,4	44,9	38,5
DA (%)	39,0	38,3	38,4	39,2	39,8	36,7
DQ (%)	11,4	7,9	15,6	19,3	15,2	24,7

EG: Exatidão global (percentagem de acertos); DA: Discordância de alocação; DQ: Discordância de quantidade.

Tabela 10. Indicadores de acurácia do melhor mapa predito e média de todos os mapas preditos com classes taxonômicas em nível de subordem na área total mapeada, dentro da área de treinamento e fora da área de treinamento.

Indicador	Melhor predição			Média geral (160 mapas)		
	Área total	Área de treinamento	Área não amostrada	Área total	Área de treinamento	Área não amostrada
EG (%)	34,3	36,9	32,1	27,6	32,9	23,1
DA (%)	52,0	52,1	46,9	48,4	45,7	44,4
DQ (%)	13,6	11,1	21,0	23,3	20,9	31,7

EG: Exatidão global (percentagem de acertos); DA: Discordância de alocação; DQ: Discordância de quantidade.

Em relação a outros estudos, constataram-se tanto valores superiores quanto inferiores para a EG na área total mapeada. Lagacherie et al. (1995), no mesmo intervalo de densidades de amostragem utilizado neste estudo, obtiveram EG entre 60% e 90% na predição de classes de solo com regras construídas a partir de áreas de referência. Grinand et al. (2008) obtiveram EG entre 59% e 67%, e verificaram resposta significativa à densidade de amostragem. Höfig et al. (2014) encontraram EG de 75% na predição de UM combinadas com árvores de decisão a partir de mapas existentes na vizinhança. Beckett & Burrough (1971), por outro lado, encontraram EG média de 53% na predição de classes de solo a partir de áreas de referência, e destacaram que esse valor se aproxima do constatado na prática em mapas de séries de solos nos EUA, que é em torno de 50%. Böhm et al. (2009) obtiveram percentagem de acertos relativamente baixa, de 26%, na extrapolação de classes de hidromorfismo.

O desempenho da predição fora da área de treinamento foi inferior ao verificado dentro da mesma, em todos os mapas preditos nos dois níveis de

classificação taxonômica dos solos, a 5% de significância. A razão entre a percentagem de acertos para classes em nível de ordem fora e dentro da área de treinamento (Tabela 9) foi de aproximadamente 85%, tanto no melhor mapa predito quanto na média de todos os 120 mapas obtidos neste nível de classificação taxonômica. Razão equivalente foi verificada no melhor mapa de classes preditas em nível de subordem (Tabela 10), porém na média dos 120 mapas obtidos neste nível de classificação a razão entre a EG fora e dentro da área de treinamento caiu para 70%. Esses dados indicam que o impacto do esquema e da densidade de amostragem, da resolução espacial ou das covariáveis utilizadas torna-se mais pronunciado à medida que se avança no nível de classificação taxonômica dos solos na área de estudo.

A razão entre o desempenho da predição fora e dentro da área de treinamento foi mais alta que a verificada em estudos similares enfocando a extrapolação. Grinand et al. (2008), por exemplo, obtiveram uma razão média entre a EG nas duas áreas de apenas 47,5%, na predição de classes de solo com árvores de decisão a partir de áreas de referência. Lemercier et al. (2012) obtiveram uma razão de 67% na predição de material de origem e de 74% na predição de classes de drenagem, com abordagem semelhante. Cambule et al. (2013), por outro lado, encontraram EG fora da área de treinamento cerca de 28% superior à da área de treinamento, na predição de teor de carbono orgânico com árvores de decisão. Entretanto, o resultado foi considerado inesperado e surpreendente, e os autores não encontraram explicação clara nem uma hipótese plausível sobre as causas.

A razão obtida neste estudo entre a área não amostrada e a área de treinamento indica que o mapa convencional de solos do Vale dos Vinhedos pode ser usado como área de referência para a predição de ordens e subordens de solos em áreas não visitadas nas proximidades. Convém, entretanto, que tais áreas sejam topograficamente semelhantes à área de referência (Scull et al., 2005; Höfig et al., 2014). A redução esperada na acurácia da predição, em ambos os níveis de classificação taxonômica, é em torno de 15% em relação à da área de treinamento, perda que pode ser considerada satisfatória tendo em vista que o processo não envolve custos adicionais com novos levantamentos. Do ponto de vista da aplicação prática, a amostragem aleatória estratificada com densidade de 0,5 pontos.ha⁻¹ e a

resolução espacial de 5 m são a melhor alternativa para a construção de árvores de decisão com este objetivo. De acordo com os resultados, essas opções têm a vantagem de propiciar simultaneamente maior magnitude e estabilidade na EG fora da área de treinamento, tanto para ordens (Figura 18b) quanto para subordens (Figura 19b).

Quanto à acurácia do mapa convencional em questão, na comparação com *pedons* independentes coletados em campo a percentagem de acertos (EG) para o nível de ordem foi de 52,4% (Tabela 11). A maior parte das discordâncias observadas entre o mapa e os *pedons* é de alocação (DA), restando uma parcela reduzida de discordância de quantidade (DQ). Para classes preditas em nível de subordem a EG foi de 29,4% (Tabela 11), sendo que neste nível de classificação a maior parte das discordâncias é relativa à DQ, enquanto a DA passou a representar uma proporção menor que no nível anterior. Isso significa que a superfície de várias subordens pode ter sido sub ou superestimada em relação à realidade durante a execução do mapeamento.

Tabela 11. Indicadores de acurácia do mapa convencional de solos do Vale dos Vinhedos para classes taxonômicas em nível de ordem e subordem, determinados a partir de 130 *pedons* independentes coletados em campo.

Indicador	Ordem	Subordem
EG (%)	52,4	29,4
DA (%)	41,3	30,1
DQ (%)	6,4	40,5

EG: Exatidão global (percentagem de acertos); DA: Discordância de alocação; DQ: Discordância de quantidade.

Uma análise visual do mapa convencional de solos com a localização dos *pedons* independentes usados para sua validação evidenciou parte das causas das discordâncias, bem como o impacto das mesmas em cada nível de classificação taxonômica (Figura 20). Por exemplo, em nível de ordem, três dos quatro pontos da Figura 20 concordaram com as classes das unidades de mapeamento (UM), o que resultaria em EG de 75% para esse nível de classificação caso somente essa área fosse avaliada. Entre as discordâncias, DA seria 0% e DQ 25%, segundo Pontius & Millones (2011). Em nível de subordem, nenhum ponto concordou com a classe das UM, de modo que a EG cairia para 0%. Dos dois pontos de Argissolo Vermelho na porção superior da Figura 20, um foi discordante devido ao deslocamento posicional

no delineamento da respectiva classe na transição entre UM adjacentes, enquanto no outro a classe efetivamente não foi delineada. O ponto de Cambissolo Háplico nas proximidades mostra outra situação de não delineamento, enquanto o de Argissolo Vermelho-Amarelo na extremidade inferior direita retrata nova discordância posicional entre UM adjacentes. Neste caso, as discordâncias seriam de 25% para DA e 75% para DQ.

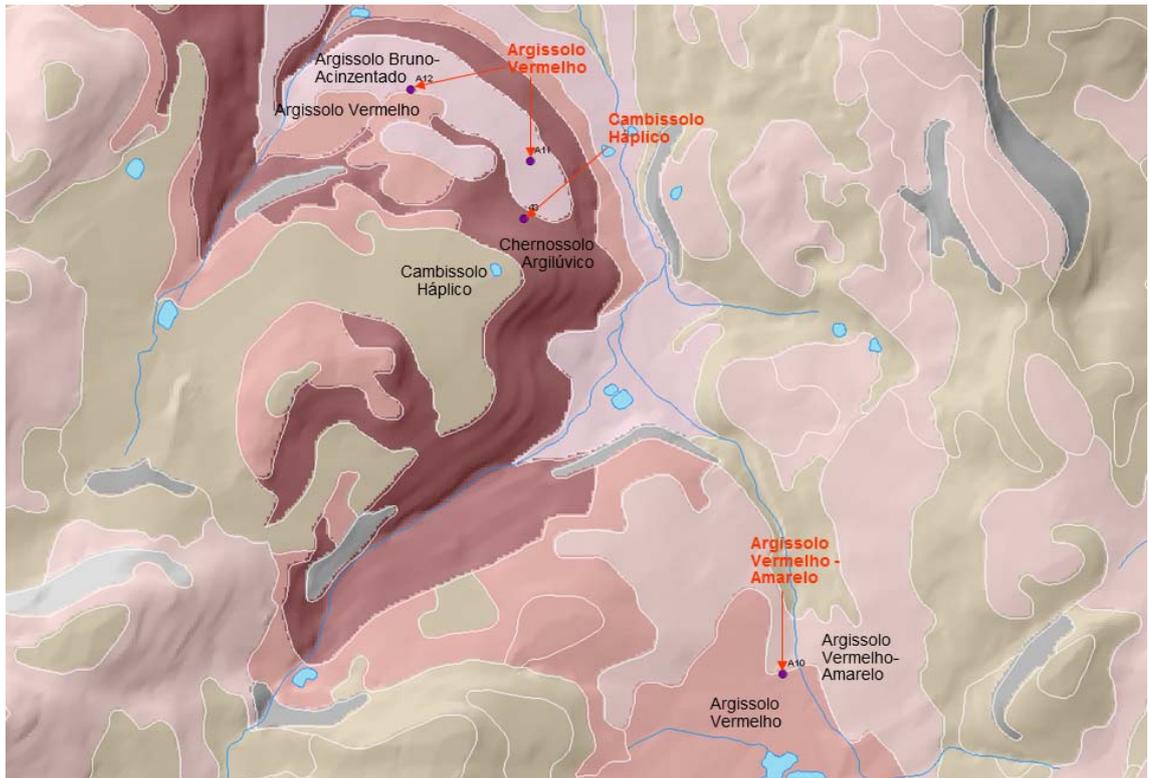
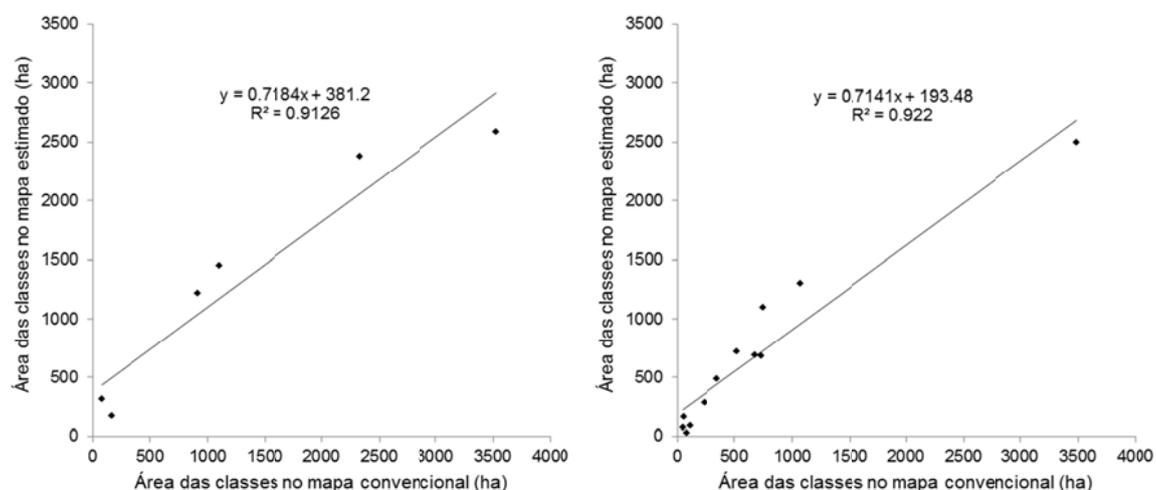


Figura 20. Exemplos de discordâncias observadas entre as classes de solo no mapa convencional e em pontos visitados em campo com distribuição espacial aleatória (o texto em vermelho corresponde à classe identificada em campo em cada ponto; o texto em preto corresponde à classe das UM nas imediações dos pontos).

A EG relativamente baixa verificada para o mapa convencional nos dois níveis de classificação contraria o senso comum, pois mapas de solos de levantamentos detalhados, constituídos de UM simples, presumidamente têm acurácia elevada. Segundo IBGE (2007), “No mínimo 70% dos *pedons* em cada delineação de uma unidade simples devem pertencer à classe taxonômica que lhe dá o nome.” Todavia, vários estudos demonstraram que a EG de mapas convencionais pode variar consideravelmente, dependendo da metodologia e propósito do levantamento e da escala de mapeamento, bem como do nível de classificação taxonômica adotado na validação. Zhu (2000), por exemplo, obteve EG de 61%, Shi et al. (2004) encontraram EG de 46%,

Chagas et al. (2007) obtiveram EG de 52,77% e Nauman & Thompson (2014), de 39%. Em geral, a validação de mapas convencionais com pontos independentes coletados em campo tende a acentuar discordâncias devido à incerteza posicional inerente ao traçado manual dos limites das UM (Rossiter, 2000; Miller, 2012; Phillips, 2013).

A percentagem de acertos (EG) obtida para os mapas preditos nos dois níveis de classificação taxonômica (Tabela 9 e Tabela 10) foi próxima da verificada para o mapa convencional usado no treinamento do algoritmo de classificação (Tabela 11). Resultado semelhante também foi verificado na comparação da área das classes de solos calculada para os respectivos mapas, nos dois níveis de classificação considerados (Figura 21). Da mesma forma, Fawcet et al. (1993) verificaram percentagem de acerto equivalente nos mapas preditos e no mapa da área de referência, e concluíram que o método de extrapolação utilizado não diferia estatisticamente do mapeamento convencional.



(a)

(b)

Figura 21. Relação entre a área total das classes de solo no mapa convencional e no melhor mapa predito, considerando a classificação taxonômica em nível de (a) ordem; e (b) de subordem.

Esses resultados indicam que o nível de acurácia esperado na predição é determinado em grande parte pela qualidade das informações de solos usadas para desenvolver as regras de classificação. Obviamente, a metodologia utilizada tem grande importância, mas em qualquer modelo não é razoável esperar resultados melhores que os dados de entrada. Quando isso

ocorre, a adequação do próprio modelo torna-se questionável, como constatado por Cambule et al. (2013).

Os resultados da validação do mapa convencional (Tabela 11) mostram que o uso de dados legados de solos para elaborar modelos de predição em MDS exige cautela, sendo recomendável uma avaliação prévia para evitar transtornos (Carré et al., 2007). Sarmiento et al. (2014) constataram que o mapa convencional utilizado neste estudo contém somente UM simples, as quais teoricamente delimitam apenas uma classe, o que é desejável para uma área de referência. No entanto, isso não é suficiente para garantir resultados satisfatórios em MDS, pois resta um grau de incerteza considerável inerente ao próprio método de mapeamento. Dessa forma, assim como ilustrado na Figura 20, pontos amostrais para treinamento podem cair em locais pouco representativos nos polígonos, introduzindo contradições e conflitos nas regras de classificação que terminam por limitar a acurácia da predição (Zhu, 2000; Grimm & Behrens, 2010).

Nesse contexto, a frequência relativa da classe predita em cada *pixel* fornece uma medida adicional para descrever a distribuição espacial da incerteza em um determinado mapa. Ela mostra como a confiança pode variar *pixel a pixel*, complementando as medidas de acurácia derivadas das matrizes de erro (dados não apresentados), que descrevem classes ou mapas como um todo. Para os mapas preditos com maior EG em nível de ordem (Figura 22) e de subordem (Figura 23), em uma parcela razoável da área de estudo as classes preditas persistiram, mesmo mudando o esquema e densidade de amostragem e o número e combinação de covariáveis explanatórias. A confiança média de toda a área para o nível de ordem foi de 0,59, com desvio padrão de 0,29, e para subordem a média foi de 0,44, com desvio padrão de 0,26. Em geral, a porção superior e mais plana e a porção inferior e mais acidentada apresentaram maior confiança, enquanto na porção intermediária do relevo, onde a ocorrência dos solos é mais diversificada e intrincada, a confiança foi menor.

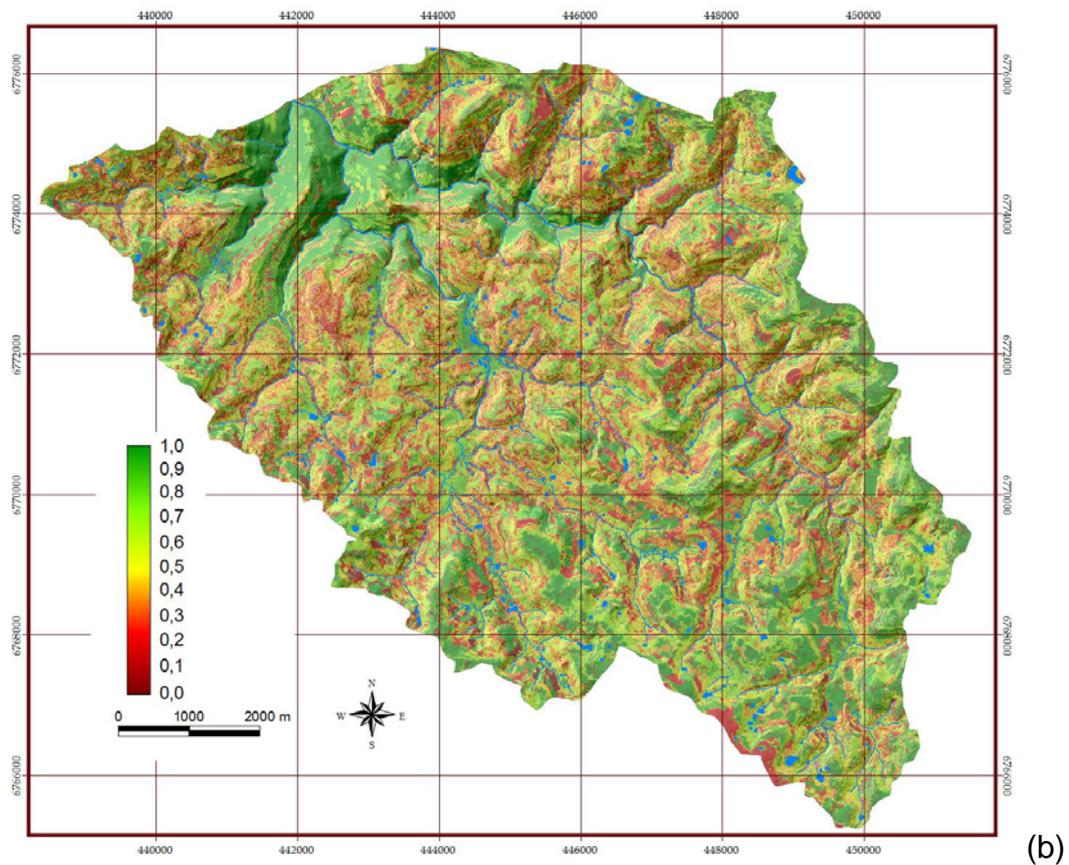
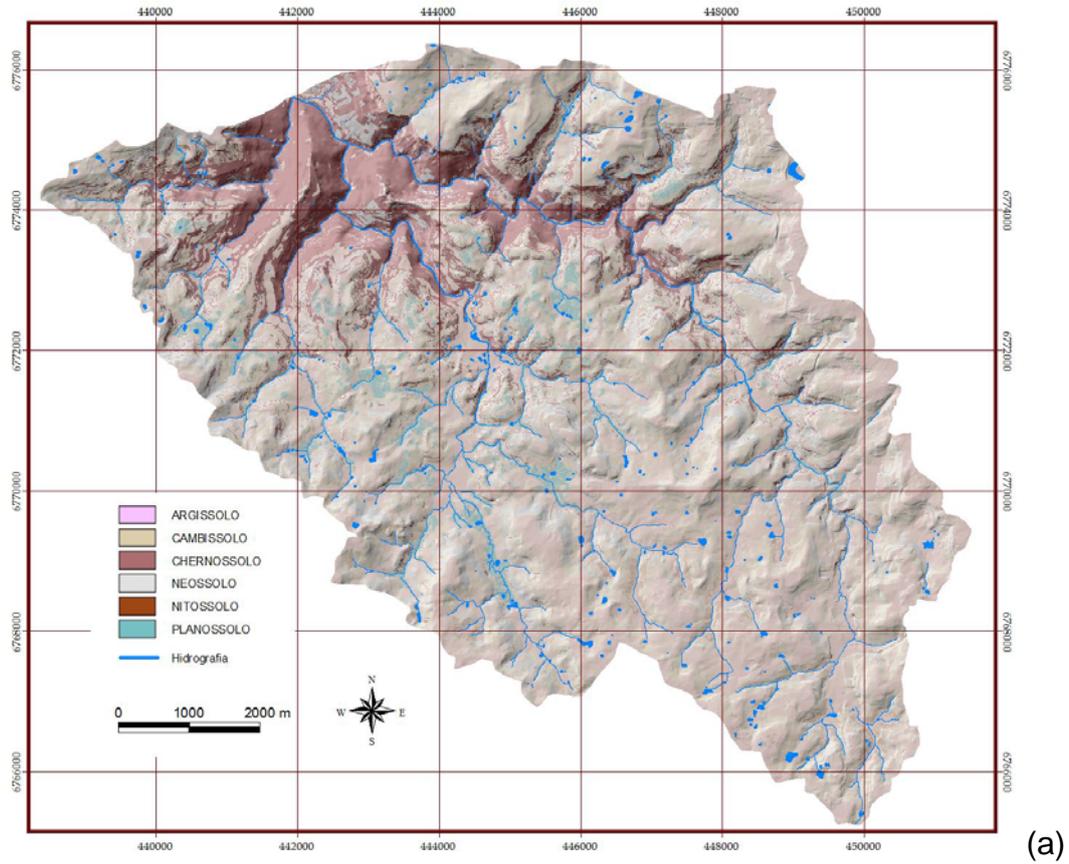


Figura 22. (a) Melhor mapa de classes previstas em nível de ordem e; (b) confiança da predição em cada *pixel*.

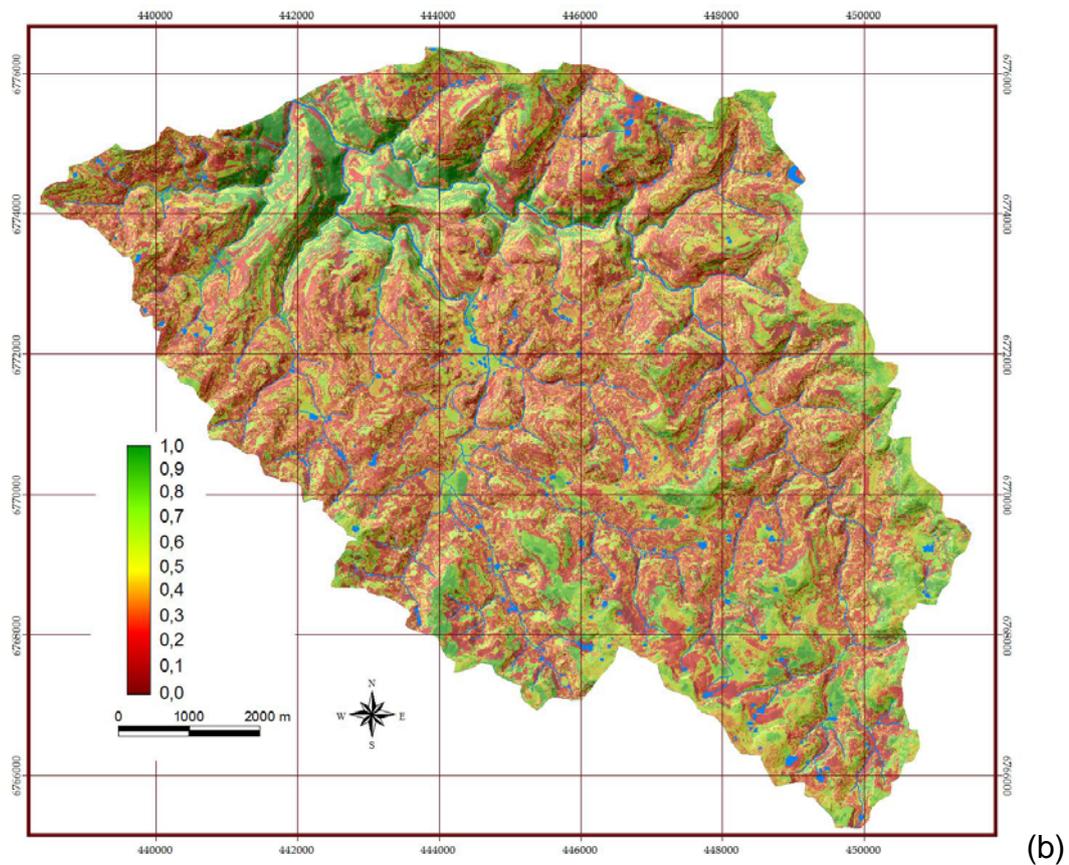
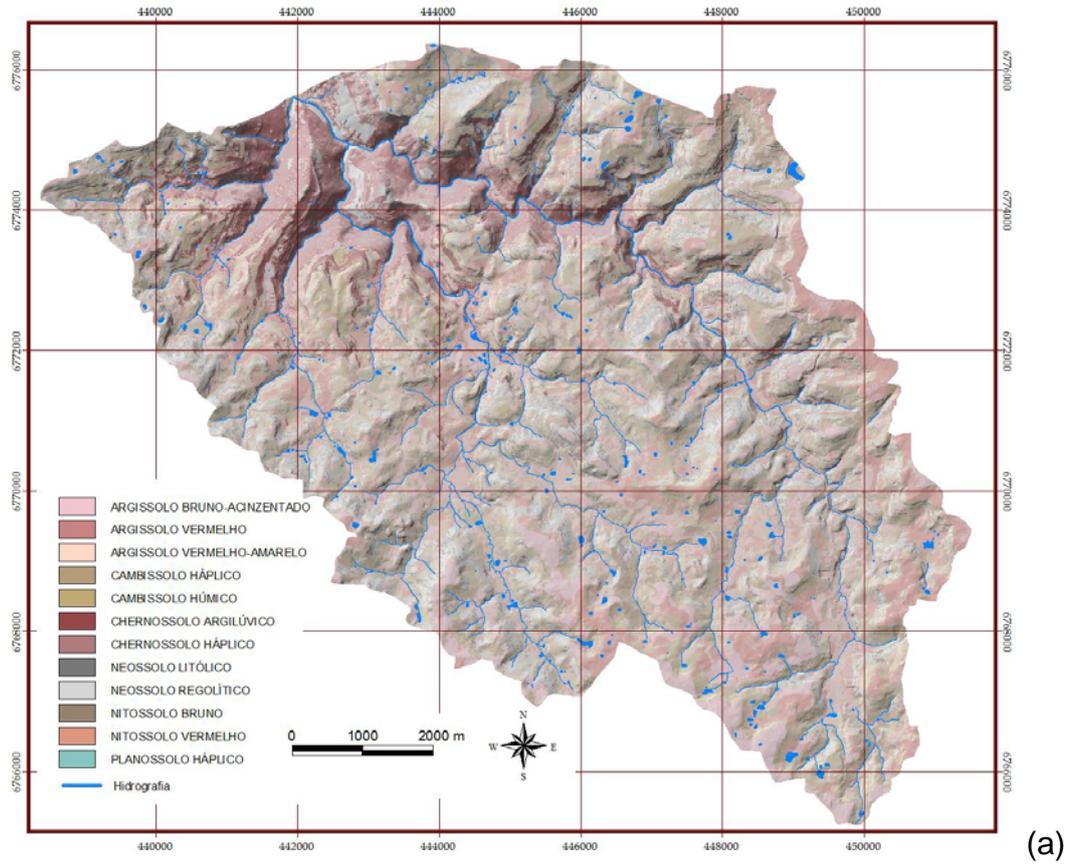


Figura 23. (a) Melhor mapa de classes previstas em nível de subordem e; (b) confiança da predição em cada *pixel*.

Em ambos os níveis de classificação taxonômica, os valores mais baixos de confiança correspondem principalmente a classes que ocupam menor superfície (Tabela 12 e Tabela 13), as quais também tenderam a apresentar erros de inclusão e de omissão mais elevados nas matrizes de erro (dados não apresentados). Em parte isso é explicado pelo menor número de pontos amostrais de treinamento obtidos nessas classes, consequência de sua área reduzida, o que eventualmente prejudicou a consistência das regras de classificação. Todavia, mesmo valores relativamente baixos, tais como 0,2 ou 0,3, por exemplo, são satisfatórios, pois significam que a classe persistiu no *pixel* em 20% ou 30% dos casos, mesmo com os parâmetros do modelo de predição sendo modificados (Bui & Moran, 2001; Nauman & Thompson, 2014).

Apesar de ser uma aproximação empírica, a frequência relativa da classe predita em cada *pixel* é útil para orientar a tomada de decisão por parte do usuário final dos mapas de classes. Ela ilustra como a confiabilidade da predição tende a variar no espaço, possibilitando assim avaliar o risco na tomada de decisão sobre o uso do solo em um determinado local (Omuto et al., 2013). Além disso, ela fornece uma medida que pode servir de base para estimar a incerteza durante a construção de novos modelos quantitativos de MDS para estimar propriedades do solo a partir das classes anteriormente preditas (Nauman & Thompson, 2014).

Tabela 12. Área ocupada e estatísticas de confiança computadas para o melhor mapa de classes preditas em nível de ordem.

Ordem	Área (ha)	Confiança			
		Mínimo	Máximo	Média	Desvio padrão
Argissolo	2.376,99	0,03	1,00	0,69	0,24
Cambissolo	2.588,64	0,03	1,00	0,48	0,26
Chernossolo	1.221,13	0,03	1,00	0,69	0,31
Neossolo	1.448,15	0,03	1,00	0,61	0,30
Nitossolo	170,03	0,03	1,00	0,40	0,20
Planossolo	316,65	0,05	1,00	0,35	0,24
Total	8.121,59				

Tabela 13. Área ocupada e estatísticas de confiança computadas para o melhor mapa de classes preditas em nível de subordem.

Subordem	Área (ha)	Confiança			
		Mínimo	Máximo	Média	Desvio padrão
Argissolo Bruno-Acinzentado	724,19	0,03	1,00	0,50	0,27
Argissolo Vermelho	1.298,58	0,03	1,00	0,45	0,23
Argissolo Vermelho-Amarelo	681,87	0,03	1,00	0,39	0,24
Cambissolo Háptico	2.494,91	0,01	1,00	0,41	0,22
Cambissolo Húmico	76,43	0,03	0,81	0,16	0,13
Chernossolo Argilúvico	287,80	0,03	1,00	0,54	0,28
Chernossolo Háptico	688,49	0,03	1,00	0,63	0,32
Neossolo Litólico	488,20	0,03	1,00	0,35	0,27
Neossolo Regolítico	1.097,76	0,01	1,00	0,42	0,25
Nitossolo Bruno	161,45	0,05	0,75	0,22	0,14
Nitossolo Vermelho	88,63	0,04	1,00	0,42	0,25
Planossolo Háptico	33,29	0,05	0,96	0,46	0,17
Total	8.121,59				

5.3.1. Considerações finais

Os resultados mostraram que o mapa convencional de solos do Vale dos Vinhedos pode ser utilizado como área de referência para a predição de classes de solo na vizinhança, desde que exista base cartográfica com detalhe equivalente. A perda de acurácia na extrapolação das classes para áreas vizinhas pode ser considerada irrelevante, porém o patamar esperado de acurácia é limitado pela qualidade do mapa existente. Isso demonstra que, se o conhecimento implícito no mapa existente pode ser recuperado e reaplicado, os erros nele contidos também o são. O nível de classificação taxonômica de ordem e subordem possibilita atender várias demandas para uso prático na região, especialmente em função da alta resolução espacial dos dados, mas convém desenvolver estudos para tentar aprofundá-lo. Nesse sentido, a predição das classes nos dois níveis considerados pode servir de base para planejar a coleta de amostras com vistas a atender esse objetivo nas duas áreas de interesse (Scull et al., 2005).

Todos os mapas resultantes da predição foram espacialmente mais detalhados que o mapa convencional usado para treinamento, o que sugere que a metodologia poderia ser aplicada também para melhorá-lo. Essa possibilidade torna-se importante frente à EG relativamente baixa verificada para o mapa existente, grande parte da qual resultante das discordâncias devido à incerteza posicional nos limites das UM. Além disso, Sarmiento et al. (2014) constataram que sua escala cartográfica efetiva (1:17.550) é um pouco inferior à nominal (1:10.000). Isso decorre do delineamento de alguns polígonos maiores que o ideal para a escala de apresentação, nos quais podem ter sido inadvertidamente incluídas outras classes além das presumidas.

O incremento no número de pontos amostrais e de covariáveis explanatórias não colaborou para elevar significativamente a EG na predição, o que concorda em parte com o verificado por Sarmiento et al. (2012) em uma porção da área de estudo. É também um aspecto positivo, pois implica em menor demanda de processamento e menor esforço de coleta em campo nas novas áreas, caso seja necessário. O esquema de distribuição espacial das amostras, por outro lado, provou ser crítico, mostrando que a localização apropriada na área estudada pode ser mais importante que a quantidade (Schmidt et al., 2008). A amostragem aleatória estratificada, associada à menor densidade de amostragem e à maior resolução espacial, possibilitou os melhores resultados. Entretanto, considerada apenas a resolução espacial, a diferença entre a de 5 m e 10 m não foi estatisticamente significativa, o que concorda parcialmente com o verificado por Cavazzi et al. (2013) e Samuel-Rosa et al. (2015).

A construção de várias árvores de decisão, partindo da premissa que a repetição de vários modelos possibilita escolher um melhor, foi semelhante ao efetuado por Bui & Moran (2001), Wei et al. (2010) e Nauman & Thompson (2014). A diferença é que a execução de todas as etapas foi conduzida manualmente e a identificação da melhor predição baseou-se na EG dos mapas obtidos, enquanto esses estudos usaram recursos de automação e elaboraram um resultado combinado a partir da classe mais frequente em cada *pixel*. Esta alternativa foi testada, mas não foi possível aplicá-la porque houve classes com frequência idêntica no mesmo *pixel*. Em nível de ordem, por

exemplo, 71% dos *pixels* apresentaram uma classe mais frequente, mas 21% apresentaram duas classes com igual frequência e 8% tiveram 3 ou mais classes nessa situação. Em nível de subordem, 59% dos *pixels* apresentaram uma classe mais frequente, 20% tiveram duas classes empatadas e 21% tiveram 3 ou mais classes empatadas. A elaboração de um resultado combinado implicaria na adoção de algum critério de desempate das classes com igual frequência.

O procedimento manual utilizado na construção das diversas árvores foi efetivo e mostrou-se apropriado para o objetivo em questão, porém foi demasiadamente trabalhoso. Em estudos futuros, convém que essa abordagem seja substituída por alternativas com maior grau de automação. É recomendável também considerar a execução de uma etapa de pré-processamento a fim de definir as covariáveis mais relevantes *a priori*, bem como selecionar áreas mais homogêneas nas UM do mapa convencional para gerar pontos amostrais mais representativos das classes de interesse. Isso pode ajudar a reduzir o ruído nos dados e a introdução de incerteza devido às inconsistências no delineamento das UM do mapa convencional e, dessa forma, elevar a acurácia dos mapas preditos a um patamar acima do verificado na área de referência (Qi & Zhu, 2003; Grinand et al., 2008; Ten Caten, 2011; Bodaghabadi et al., 2011).

5.4. Conclusões

- i. A metodologia utilizada possibilita a predição de classes de solo em nível de ordem e de subordem em áreas não visitadas, com alta resolução espacial, com acurácia equivalente à do mapa convencional usado para treinamento do algoritmo de classificação;
- ii. Na região estudada, a predição pode ser estendida para áreas não visitadas com características topográficas semelhantes e extensão até 1,5 vezes maior que a área de treinamento. A expansão para áreas maiores precisa ser testada;
- iii. O esquema de distribuição espacial e a densidade de amostragem na geração dos pontos para treinamento são determinantes para o grau de acurácia da predição, sendo recomendável a amostragem aleatória estratificada e densidade de 0,5 pontos.ha⁻¹. Densidades menores precisam ser testadas.

6. CONCLUSÕES GERAIS

Os resultados indicam que os dados legados de solos existentes no Brasil ainda têm um papel importante a desempenhar na produção de mais e melhores mapas de solos, pois podem fornecer muito mais informações do que aparentam conter à primeira vista. A desagregação oferece uma possibilidade de incrementar o detalhe dos mapas existentes e de harmonizar e compatibilizar mapas produzidos por diferentes levantamentos, a baixíssimo custo. O uso de conhecimento formal sobre as relações solo-paisagem para individualizar as classes de unidades de mapeamento combinadas tem a vantagem de contornar a escassez de dados analíticos e descritivos sobre os solos amparando-se nas bases conceituais da pedologia.

A predição em áreas não visitadas a partir de áreas de referência, por sua vez, oferece a oportunidade de expandir a cobertura dos mapeamentos a partir de mapas existentes. Essa abordagem de extrapolação é especialmente importante no caso de mapas mais detalhados, que demandam alta intensidade de amostragem, à medida que dados espaciais com alta resolução espacial estão se tornando disponíveis e podem fornecer covariáveis preditoras apropriadas no futuro próximo. Ambas as estratégias de MDS também podem ser usadas complementarmente: mapas de classes individualizadas por desagregação de unidades de mapeamento, por exemplo, podem ser usadas como treinamento para predizer classes de solo em áreas onde ainda não existem mapas.

Em ambos os casos estudados, a qualidade dos dados legados de solos foi determinante para a acurácia nos resultados das predições, o que demonstra a necessidade premente de levá-la em conta ao desenvolver e calibrar modelos de predição em MDS. Isso é ainda mais relevante frente ao fato de que a qualidade de mapas convencionais de solo, por uma série de razões, tende a ser inferior à presumida, o que não é exclusividade do Brasil. Caso a qualidade dos dados existentes não seja considerada de alguma forma,

corre-se o risco de superestimar inadvertidamente a importância dos métodos e/ou das covariáveis sobre a acurácia dos resultados e, conseqüentemente, de formular conclusões ou hipóteses equivocadas.

7. CONSIDERAÇÕES FINAIS

Os estudos que compõem esta tese demonstram o potencial e a viabilidade de algumas estratégias e métodos de mapeamento para detalhar e expandir a cobertura de mapas de classes de solos. Embora sejam promissores, eles se referem a áreas de estudo limitadas e requerem a realização de novos estudos a fim de confirmar e generalizar hipóteses e resultados, bem como aperfeiçoar as metodologias propostas.

Com relação à desagregação embasada no conhecimento formal das relações solo-paisagem, convém testar a metodologia em áreas com características de relevo distintas e/ou com disponibilidade de mapas convencionais em outras escalas. Da mesma forma, é importante avaliar o efeito da definição dos parâmetros de classificação no mapeamento de classes de geformas com o programa LandMapr, bem como avaliar outras alternativas para individualizar classes de formas de relevo.

Quanto à predição a partir de áreas de referência, é recomendável testar a aplicação de múltiplos modelos de árvores de decisão usando métodos automatizados, tais como o emprego de florestas randômicas (*random forests*). No caso específico da região estudada, também é importante validar o uso dos dados detalhados disponíveis no Vale dos Vinhedos para predizer classes em áreas que ainda não foram efetivamente mapeadas, para as quais existem bases cartográficas com nível de detalhe equivalente.

Em ambos os casos, convém avaliar a possibilidade de uso dos resultados da predição de classes para elaborar mapas de algumas propriedades dos solos. A importância de dados de campo para a validação de modelos de predição a partir de dados legados, tanto de classes quanto de propriedades de solos, também torna importante considerar estudos futuros voltados a novos métodos de aquisição usando técnicas pedométricas.

8. REFERÊNCIAS

AHRENS, R. J.; ARNOLD, R. W. Soil Taxonomy. In: HUANG, P. M.; LI, Y.; SUMNER, M. E. **Handbook of Soil Sciences: properties and processes**. Boca Raton: Taylor & Francis, 2012, p. 31-1–31-33.

ARNOLD, R. W. Concepts of Soils and Pedology. In: WILDING, L.P.; SMECK, N.E.; HALL, G.F. (Ed.). **Pedogenesis and Soil Taxonomy**. I. Concepts and Interactions. Amsterdam: Elsevier, 1983, p. 1–21.

ASHTEKAR, J. M.; OWENS, P. R. Remembering Knowledge: An Expert Knowledge Based Approach to Digital Soil Mapping. **Soil Horizons**, Madison, 2014, 6 p.

BARRINGER, J. et al. National Mapping of Landform Elements in Support of S-Map, A New Zealand Soils Database. In: ZHOU, Q.; LEES, B.; TANG, G. (Ed.). **Advances in Digital Terrain Analysis**. Berlin: Springer, 2008. p. 443-458.

BASAYIGIT, L.; SENOL, S. Comparison of soil maps with different scales and details belonging to the same area. **Soil & Water Research**, Amsterdam, v. 3, p. 31–39, 2008.

BECKETT, P. H. T.; BURROUGH, P. A. The relation between cost and utility of soil surveys: IV: comparison of the utilities of soil maps produced by different survey procedures, and at different scales. **Journal of Soil Science**, Oxford, v. 22, p. 465–480, 1971.

BIE, S. W.; BECKETT, P. H. T. Comparison of four independent soil surveys by air-photo interpretation, Paphos area (Cyprus). **Photogrammetria**, Amsterdam, v. 29, p. 189-202, 1973.

BOCKHEIM, J. G, et al. Historical development of key concepts in pedology. **Geoderma**, Amsterdam, v. 124, p. 23–36, 2005.

BODAGHABADI, M. B. et al. Using Canonical Correspondence Analysis (CCA) to identify the most important DEM attributes for digital soil mapping applications. **Catena**, Amsterdam, v. 86, p. 66–74, 2011.

BOETTINGER, J. L. Environmental Covariates for Digital Soil Mapping in the Western USA. In: BOETTINGER, J. L. et al. (Ed.). **Digital Soil Mapping: bridging research, environmental application, and operation**. London: Springer, 2010, p. 17-27.

BÖHM, C. et al. An approach to extrapolate categorical agricultural soil data to nonmapped areas using majority vote. **Journal of Plant Nutrition and Soil Science**, Weinheim, v. 172, p. 467–476, 2009.

BRAGATO, G. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. **Geoderma**, Amsterdam, v. 118, p. 1–16, 2004.

BROWN, D. J. A Historical Perspective on Soil-Landscape Modeling. In: GRUNWALD, S. (Ed.). **Environmental soil-landscape modeling: geographic information technologies and pedometrics**. Boca Raton: Taylor & Francis, 2006. p. 62–103.

BUI, E. N.; MORAN, C. J. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. **Geoderma**, Amsterdam, v. 103, p. 79–94, 2001.

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

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

BUOL, S. W. et al. **Soil Genesis and Classification**. Chichester: Wiley-Blackwell, 2011. 543 p.

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**, Netherlands, v. 113, p. 37–52, 2000.

CAMBULE, A. H.; ROSSITER, D. G.; STOOORVOGEL, J. J. A methodology for digital soil mapping in poorly-accessible áreas. **Geoderma**, Amsterdam, v. 192, p. 341–353, 2013.

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**, Bethesda, v. 55, p. 1613–1618, 1989.

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

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

CHAGAS, C. S. et al. Utilização de redes neurais artificiais para predição de classes de solo em uma bacia hidrográfica no Domínio do Mar de Morros. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., Florianópolis, 2007. **Anais...** Florianópolis, INPE, 2007. p. 2421–2428.

CHIVERRELL, R. C., THOMAS, G. S. P.; FOSTER, G. C. Sediment–landform assemblages and digital elevation data: Testing an improved methodology for the assessment of sand and gravel aggregate resources in north-western Britain. **Engineering Geology**, Amsterdam, v. 99, p. 40–50, 2008.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, Thousand Oaks, 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**, New York, v. 37, p. 35–46, 1991.

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

D'AVELLO, T. P. & McLEESE, R. L. Why are those lines placed where they are? An investigation of soil map recompilation methods. **Soil Survey Horizons**, Madison, v. 39, p. 119-126, 1998.

DEBELLA-GILO, M.; ETZELMÜLLER, B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. **Catena**, Amsterdam, v. 77, p. 8–18, 2009.

DE BRUIN, S., WIELEMAKER, W. G., MOLENAAR, M. Formalisation of soil-landscape knowledge through interactive hierarchical disaggregation. **Geoderma**, Amsterdam, v. 91 (1–2), p. 151–172, 1999.

DE REU, J. et al. Application of the topographic position index to heterogeneous landscapes. **Geomorphology**, Amsterdam, v. 186, p. 39–49, 2013.

EASTMAN, J. R. **IDRISI Selva – GIS and Image Processing Software – version 17.0**. Worcester: Clark Labs, 2012.

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**, Basel, v. 2, p. 151–165, 2010.

EMBRAPA. Centro Nacional de Pesquisa de Solos. **Sistema Brasileiro de Classificação de Solos**. 3. ed. Brasília, 2013. 353 p.

EMBRAPA UVA E VINHO. **Normal climatológica. Estação Agroclimática da Embrapa Uva e Vinho, Bento Gonçalves, RS. Período de 1961 a 1990**. Bento Gonçalves: Embrapa Uva e Vinho. 2008. Disponível em: <<http://www.cnpuv.embrapa.br/>>. Acesso em: 05 nov. 2010.

ESPÍNDOLA, C. R. A pedologia e a evolução das paisagens. **Revista do Instituto Geológico**, São Paulo, v. 31, n. 1/2, p. 67–92, 2010.

ESRI. **ArcGIS10.0**. Redlands: Environmental Systems Research Institute. 2013.

EVANS, I. S. Geomorphometry and landform mapping: What is a landform? **Geomorphology**, Amsterdam, v. 137, p. 94–106, 2012.

FAWCET, M. D. et al. **An evaluation of the extrapolatory method of soil soil mapping**. Edmonton: Environmental and Engineering Department, Alberta Research Council. 1993. 69p. (Alberta Research Council Open File Report 1993–09). Disponível em: <http://www.ag.gov.ab.ca/publications/OFR/PDF/OFR_1993_09.PDF>. Acesso em: 15 mar. 2014.

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

FINNEN, E.; MENZA, C. **Sampling Design Tool for ArcGIS**. Silver Spring: NOAA/NOS/Center for Coastal Monitoring and Assessment. Disponível em: <coastalscience.noaa.gov/products/tools>. Acesso em: 29 dez. 2013.

FISK, H. et al. Department of Agriculture (USDA) TEUI Geospatial Toolkit: An Operational Ecosystem Inventory Application. In: BOETTINGER, J. L. et al. (Ed.). **Digital Soil Mapping: bridging research, environmental application, and operation**. London: Springer, 2010. p. 399–410.

FLORES, C. A. (Org.) et al. **Levantamento semidetalhado de solos: região da campanha – Folha Palomas, Estado do Rio Grande do Sul**. Porto Alegre: Editora da UFRGS, 2007a. 95 p.

FLORES, C. A. et al. Levantamento semidetalhado dos solos na Serra do Sudeste, RS como subsídio ao zoneamento vitivinícola: Folha Pinheiro Machado. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 31., Gramado, 2007. **Anais...** Gramado: Sociedade Brasileira de Ciência do Solo, 2007b.

FLORES, C. A. et al. **Levantamento semidetalhado de solos: Região da Serra Gaúcha - Rio Grande do Sul**. Porto Alegre: UFRGS/Embrapa Clima Temperado, 2007c.

FLORES, C. A. et al. Potencial edáfico da Serra Gaúcha, Brasil para viticultura. In: CONGRESSO LATINOAMERICANO DE VITICULTURA E ENOLOGIA, 11., Mendoza, 2007. **Anais...** Mendoza, IVV: CLEIFRA: SECYT, 2007d.

FLORES, C. A. et al. Levantamento semidetalhado dos solos na Serra do Sudeste, RS como subsídio ao zoneamento vitivinícola: Folha Encruzilhada do Sul. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 32., Fortaleza, 2009. **Anais...** Fortaleza: Sociedade Brasileira de Ciência do Solo, 2009. 1 CD-ROM.

FLORES, C. A. et al. **Os solos do Vale dos Vinhedos**. Brasília: Embrapa, 2012. 176 p., 2 mapas. 1 DVD-ROM.

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

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

FORBES, T. R.; ROSSITER, D.; VAN WAMBEKE, A. **Guidelines for evaluating the adequacy of soil resource inventories**. Ithaca: Cornell University Department of Agronomy, 1987. Disponível em: <<http://eprints.icrisat.ac.in/8766/1/RP%2D02147.pdf>>. Acesso em: 03 dez. 2013.

GENG, X. et al. Toward digital soil mapping in Canada: Existing soil survey data and related expert knowledge. In: BOETTINGER, J.L. et al. (Ed.). **Digital Soil Mapping: bridging research, environmental application, and operation**. London: Springer, 2010. p. 325–335.

GIASSON, E. et al. Use of decision tree analysis for predictive soils mapping on subtropical basaltic steep lands. **Scientia Agrícola**, Piracicaba, v. 68, n. 2, p. 167–174, 2011.

GOOGLE. **Google Earth**. Mountain View: Google Inc. 2014.

- ESRI. **ArcGIS10.0**. Redlands: Environmental Systems Research Institute. 2013.
- GRIMM, R.; BEHRENS, T. Uncertainty analysis of sample locations within digital soil mapping approaches. **Geoderma**, Amsterdam, v. 155, p. 154–163, 2010.
- GRINAND, C. et al. 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.
- GRUNWALD, S. (Ed.). **Environmental Soil-Landscape Modeling**: geographic information technologies and pedometrics. Boca Raton: Taylor & Francis, 2006. 488 p.
- HANSEN, M. K. et al. Inductively mapping expert-derived soil-landscape units within Dambo Wetland catenae using multispectral and topographic data. **Geoderma**, Amsterdam, v. 150, p. 72–84, 2009.
- HÄRING, T. et al. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. **Geoderma**, Amsterdam, v. 185–186, p. 37–47, 2012.
- HARTEMINK, A. E.; McBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008.
- HARTEMINK, A. E. Soil Map Density and a Nation's Wealth and Income. In: HARTEMINK, A. E.; McBRATNEY, A.; MENDONÇA-SANTOS, M. de L. (Ed.). **Digital Soil Mapping With Limited Data**. New York: Springer, 2008, p. 53–66.
- HARTEMINK, A. E.; McBRATNEY, A. A soil science renaissance. **Geoderma**, Amsterdam, v. 148, p. 123–129, 2008.
- HARTEMINK, A. E. et al. GlobalSoilMap.net: a new digital soil map of the world. In: BOETTINGER, J.L., et al. (Ed.). **Digital Soil Mapping. Bridging Research, Environmental Application, and Operation**. London: Springer, 2010, p. 423–428.
- HARTEMINK, A. E.; KRASILNIKOV, P.; BOCKHEIM, J.G. Soil maps of the World. **Geoderma**, Amsterdam, v. 207–208, p. 256–267, 2013.
- HARTEMINK, A. E.; MINASNY, B. Towards digital soil morphometrics. **Geoderma**, Amsterdam, v. 230–231, p. 305–317, 2014.
- HASENACK, H.; WEBER, E.(Org.) **Base cartográfica digital da Serra Gaúcha - escala 1:50.000**. Porto Alegre: UFRGS Centro de Ecologia. 2007. 1 CD-ROM. (Série Geoprocessamento n.2).
- HEMPEL, J. W. et al. 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.
- HENGL, T.; HUSNJAK, S. Evaluating adequacy and usability of soil maps in Croatia. **Soil Science Society of America Journal**, Madison, v. 70, p. 920–929, 2006.
- HENGL, T.; ROSSITER, D. G. Supervised Landform Classification to Enhance and Replace Photo-Interpretation in Semi-Detailed Soil Survey. **Soil Science Society of America Journal**, Madison, v. 67, p. 1810–1822, 2003.

- HENGL, T. et al. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, Amsterdam, v. 140, p. 417–427, 2007.
- HEUVELINK, G. B. M.; WEBSTER, R. Modelling soil variation: past, present, and future. **Geoderma**, Amsterdam, v. 100, p. 269–301, 2001.
- HOLE, F. D. An approach to landscape analysis with emphasis on soils. **Geoderma**, Amsterdam, v. 21, p. 1–23, 1978.
- HOLMES, K. W. et al. Spatial disaggregation of conventional soil mapping across Western Australia using DSMART. In: ARROUAYS, D. et al. (Ed.). **GlobalSoilMap: Basis of the global spatial soil information system**. London: Taylor & Francis, 2014, p. 273–279.
- HÖFIG, P.; GIASSON, E.; VENDRAME, P. R. S. Mapeamento digital de solos com base na extrapolação de mapas entre áreas fisiograficamente semelhantes. **Pesquisa Agropecuária Brasileira**, Brasília, v. 49, n. 12, p. 958–966, 2014.
- HUDSON, B. D. The Soil Survey as Paradigm-based Science. **Soil Science Society of America Journal**, Madison, v. 56, p. 836–841, 1992.
- HUNTER, G. J.; WACHOWICZ, M.; BREGT, A. K. Understanding spatial data usability. **Data Science Journal**, Paris, v. 2, p. 79–89, 2003.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Levantamento de recursos naturais (Folha SH.22 Porto Alegre e parte das Folhas SH.21 Uruguaiana e SI.22 Lagoa Mirim)**. Rio de Janeiro: IBGE, 1986. 1 CD-ROM
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Manual técnico de pedologia**. 2.ed. Rio de Janeiro: IBGE, 2007. 320p.
- JENNY, H. **Factors of Soil Formation: A System of Quantitative Pedology**. New York: McGraw-Hill. 1941. 264p.
- JOHNSON, D. L.; SCHAETZL, R. J. Differing views of soil and pedogenesis by two masters: Darwin and Dokuchaev. **Geoderma**, Amsterdam, v. 237–238, p. 176–189, 2015.
- JUNGBLUT, M. **Pedologia da bacia do rio Gravataí-RS**. Porto Alegre: CPRM/Metroplan, 1994. 29 p. (Série Cartas Temáticas, v.2).
- KELLOG, C. E. Soil Genesis, Classification, and Cartography: 1924–1974. **Geoderma**, Amsterdam, v. 12, p. 347–362, 1974.
- KERRY, R. et al. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. **Geoderma**, Amsterdam, v. 170, p. 347–358, 2012.
- KIM, D. & ZHENG, Y. Scale-dependent predictability of DEM based landform attributes for soil spatial variability in a coastal dune system. **Geoderma**, Amsterdam, v. 164, p. 181–194, 2011.
- LAGACHERIE, P. Digital Soil Mapping: A State of the Art. In.: HARTEMINK, A.E.; McBRATNEY, A.; MENDONÇA-SANTOS, M.L. (Ed.). **Digital Soil Mapping with Limited Data**. New York: Springer, 2008, p. 3-14.

LAGACHERIE, P.; LEGROS, J. P.; BURROUGH, P. A. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. **Geoderma**, Amsterdam, v. 65, p. 283–301, 1995.

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

LEGROS, J.P. **Major Soil Groups of the World: Ecology, Genesis, Properties and Classification**. Boca Raton: CRC Press, 2012. 454 p.

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

LI, S. et al. Extracting topographic characteristics of landforms typical of Canadian agricultural landscapes for agri-environmental modeling. I. Methodology. **Canadian Journal of Soil Science**, Ottawa, v. 91, p. 251–266, 2011.

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

LIN, H. et al. Assessment of soil spatial variability at multiple scales. **Ecological Modelling**, Amsterdam, v. 182, p. 271–290, 2005.

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

LORENZETTI, R. et al. Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small scale maps. **Geoderma**, Amsterdam, v. 237–238, p. 237–245, 2015.

MACHADO, M. L. L.; FITZ, P. **Rio Grande do Sul. Mapa de solos. Generalização cartográfica do levantamento de reconhecimento de solos do Rio Grande do Sul - Brasil, 1973**. Porto Alegre: Emater, 2001.

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

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

MAcMILLAN, R. A. et al. Predictive Ecosystem Mapping (PEM) for 8.2 Million ha of Forestland, British Columbia, Canada. In: BOETTINGER, J.L. et al. (Ed.). **Digital Soil Mapping: Bridging Research, Environmental Application, and Operation**. London: Springer, 2010, p. 337–356.

MALONE, B. P.; McBRATNEY, A. B.; MINASNY, B. Spatial Scaling for digital soil mapping. **Soil Science Society of America Journal**, Madison, v. 77, p. 890–902, 2012.

- MALONE, B. P. et al. Using model averaging to combine soil property rasters from legacy soil maps and from point data. **Geoderma**, Amsterdam, v. 232–234, p. 34–44, 2014.
- McBRATNEY, A. B.; MENDONÇA-SANTOS, M.L.; MINASNY, B. On Digital Soil Mapping. **Geoderma**, Amsterdam, v. 117, p. 3–52, 2003.
- McBRATNEY, A. B. et al. Digital Soil Mapping. In: HUANG, P.M.; LI, Y.; SUMNER, M.E. **Handbook of Soil Sciences: Properties and Processes**. Boca Raton: Taylor & Francis, 2012, p. 37-43.
- McBRATNEY, A. B. et al. An overview of pedometric techniques for use in soil survey. **Geoderma**, Amsterdam, v. 97, p. 293–327, 2000.
- McKENZIE, N. J.; RYAN, P. J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, Amsterdam, v. 89, p. 67–94, 1999.
- MEDEIROS, J. S.; CÂMARA, G. **Geoprocessamento para projetos ambientais**. São José dos Campos: INPE, 2001. Disponível em: <<http://mtc-m12.sid.inpe.br/col/sid.inpe.br/sergio/2004/04.19.15.08/doc/cap10-aplicacoesambientais.pdf>>. Acesso em: 06 jan. 2011.
- MENDONÇA-SANTOS, M. L.; SANTOS, H. G. The state of the art of Brazilian soil mapping and prospects for digital soil mapping. In: LAGACHERIE, P.; McBRATNEY, A.B.; VOLTZ, M. (Ed.). **Digital Soil Mapping: an introductory perspective**. Amsterdam: Elsevier, 2007. p. 39–54. (Developments in soil science, vol. 31).
- MILLER, B. A. The Need to continue improving soil survey maps. **Soil Horizons**, Madison, v. 53, p. 14–15, 2012.
- MILLER, B. A., SCHAETZL, R. J. The historical role of base maps in soil geography. **Geoderma**, Amsterdam, v. 230–231, p. 329–339, 2014.
- MINASNY, B. & McBRATNEY, A. B. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. **Geoderma**, Amsterdam, v. 142, p. 285–293, 2007.
- MINASNY, B.; McBRATNEY, A. B. Methodologies for Global Soil Mapping. In: BOETTINGER, et al. (Ed.). **Digital Soil Mapping: Bridging Research, Environmental Application, and Operation**. London: Springer, 2010, p. 429–436.
- MÖLLER, M. et al. Plausibility test of conceptual soil maps using relief parameters. **Catena**, Amsterdam, v. 88, p. 57–67, 2012.
- MOORE, I. D., GRAYSON, R. B., LADSON, A. R. Digital terrain modeling – a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, Chichester, v. 5, n.1, p. 3–30, 1991.
- MOORE, I. D. et al. Soil attribute prediction using terrain analysis. **Soil Science Society of America Journal**, Madison, v. 57, p. 443–452, 1993.
- NAUMAN, T. W. et al. Fuzzy disaggregation of conventional soil maps using database knowledge extraction to produce soil property maps. In: MINASNY, B.; MALONE; B.P.; McBRATNEY, A.B. (Ed.). **Digital Soil Assessments and Beyond**. London: Taylor & Francis. 2012.
- NAUMAN, T. W.; THOMPSON, J. A.; RASMUSSEN, C. Semi-automated disaggregation of a conventional soil map using knowledge driven data mining

and random forests in the Sonoran Desert, USA. **Photogrammetric Engineering & Remote Sensing**, Bethesda, v. 80, n. 4, p. 353–366, 2014.

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

NUSSBAUM, M. et al. The relevance of scale in soil maps. **Bulletin BGS**, [Zurich], v. 32, p. 63–70, 2011.

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

ODGERS, N. P.; McBRATNEY, A. B.; MINASNY, B. Digital soil property mapping and uncertainty estimation using soilclass probability rasters. In: ARROUAYS, D. et al. (Ed.). **GlobalSoilMap: Basis of the global spatial soil information system**. London: Taylor & Francis, 2014b. p. 341–346.

OLIVEIRA, J. B. O novo sistema brasileiro de classificação de solos. **O Agrônomo**, Campinas, v. 53, n. 1, p. 8–10, 2001.

OLSON, C. G. Geomorphological Soil-Landscape Models. In: GRUNWALD, S. (Ed.). **Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics**. Boca Raton: Taylor & Francis, 2006. p. 105–124.

OMUTO, C.; NACHTERGAELE, F.; ROJAS, R. V. **State of the art report on global and regional soil information: Where are we? Where to go?** Global Soil Partnership Technical Report. Roma: FAO, 2013. 69p. Disponível em: <<http://www.fao.org/docrep/017/i3161e/i3161e.pdf>>. Acesso em: 15 dez. 2013.

PHILLIPS, J. D. On the relations between complex systems and the factorial model of soil formation (with Discussion). **Geoderma**, Amsterdam, v. 86, p. 1–21, 1998.

PHILLIPS, J. D. Evaluating taxonomic adjacency as a source of soil map uncertainty. **European Journal of Soil Science**, Oxford, v. 64, p. 391–400, 2013.

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

PONTIUS, R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, London, v. 32, p. 4407–4429, 2011.

PROTAS, J. F. S; CAMARGO, U. A.; MELLO, L. M. R. Vitivinicultura brasileira: regiões tradicionais e pólos emergentes. **Informe Agropecuário**, Belo Horizonte, v. 27, n. 234, p. 7–15, 2006.

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

ROECKER, S. M. et al. Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches. In: BOETTINGER, J.L. et al. (Ed.).

Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. London: Springer, 2010, p. 369–384.

ROSSITER, D. G. **Methodology for soil resource inventories**. Enschede: ITC, 2000. 2.ed. 132p. Disponível em: <http://www.itc.nl/personal/rossiter/teach/ssm/SSM_LectureNotes2.pdf>. Acesso em: 13 jan. 2014.

ROSSITER, D. G. **Assessing the thematic accuracy of area-class soil maps**. Enschede: ITC, 2001. 46p. Disponível em: <http://www.itc.nl/~rossiter/docs/AccuracySoilMaps_Preprint31Aug2001.pdf> Acesso em: 20 abr. 2009.

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

ROSSITER, D. G. Digital soil mapping: Towards a multiple-use Soil Information System. In: SEMANA DE LA GEOMÁTICA, Santa Fé de Bogotá, 2005. **Anais...** Santa Fé de Bogotá, Universidad Nacional de Colombia, 2005. Disponível em: <<http://www.itc.nl/personal/rossiter>>. Acesso em: 24 set. 2009.

RUKHOVICH, D. I. et al. State Soil Map of the Russian Federation: An ArcInfo version. **Eurasian Soil Science**, Silver Springer, v. 46, p. 225-40, 2013.

SAA/IBGE-RS/SC. **Mapa de solos do Rio Grande do Sul, escala 1:250.000**. Porto Alegre: SAA-RS/IBGE-Unidade SC, 2003. (Convênio Secretaria de Agricultura e Abastecimento e Instituto Brasileiro de Geografia e Estatística/Unidade Estadual de Santa Catarina).

SAMUEL-ROSA, A. et al. Do more detailed environmental covariates deliver more accurate soil maps? **Geoderma**, Amsterdam, v. 243–244, p. 214–227, 2015.

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

SANTOS, H. H. Comissão de solos. **Boletim Informativo da SBCS**, Campinas, v. 22, n. 1, p. 10–30, 1997.

SARMENTO, E. C. et al. Sistema de informação geográfica como apoio ao levantamento detalhado de solos do Vale dos Vinhedos. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, p. 795–2803, 2008.

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

SARMENTO, E. C. et al. Avaliação da escala efetiva de dois mapas de solos da mesma área na Serra Gaúcha. In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO, 34., Florianópolis, 2013. **Anais...** Florianópolis: Sociedade Brasileira de Ciência do Solo, 2013.

SARMENTO, E. C. et al. Caracterização de mapas legados de solos: uso de indicadores em mapas com diferentes escalas no Rio Grande do Sul. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 38, p. 1672–1680, 2014.

SCHAETZL, R.; ANDERSON, S. **Soils: genesis and geomorphology**. New York: Cambridge University Press, 2005. 817 p.

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

SCHNEIDER, P. et al. Diagnóstico ambiental de Porto Alegre: solos. In: HASENACK, H.; WEBER, E.; MARCUZZO, S. (Org.) **Diagnóstico ambiental de Porto Alegre: Geologia, solos, drenagem, vegetação e ocupação**. Porto Alegre: Secretaria Municipal do Meio Ambiente, 2008. p. 28-43.

SCULL, P. et al. 2003. Predictive soil mapping: a review. **Progress in Physical Geography**, London, 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**, Amsterdam, v. 181, p. 1–15, 2005.

SHI X. et al. A Case-based Reasoning Approach to Fuzzy Soil Mapping. **Soil Science Society of America Journal**, Madison, v. 68, p. 885–894, 2004.

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

STRECK, E. V. et al. **Solos do Rio Grande do Sul**. 2.ed. Porto Alegre: EMATER/RS, 2008. 222 p.

STUM, A.K. et al. Random Forests Applied as a Soil Spatial Predictive Model in Arid Utah. In: BOETTINGER, J.L. et al. (Ed.). **Digital Soil Mapping: Bridging Research, Environmental Application, and Operation**. London: Springer, 2010, p. 179-189.

SUBBURAYALU, S. K. SLATER, B. K. Soil series mapping by knowledge discovery from an Ohio County Soil Map. **Soil Science Society of America Journal**, Madison, v. 77, p. 1254-1268, 2013.

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

THOMPSON, J.A. et al. Regional approach to soil property mapping using legacy data and spatial disaggregation techniques. In: WORLD CONGRESS OF SOIL SCIENCE, SOIL SOLUTIONS FOR A CHANGING WORLD, 19., Brisbane, 2010. **Proceedings...** Brisbane: International Union of Soil Science, 2010. 1 DVD-ROM.

TEN CATEN, A. **Mapeamento digital de solos: metodologias para atender a demanda por informação espacial em solos**. 2011. 104 f. Tese (Doutorado) – Programa de Pós-graduação em Ciência do Solo, Faculdade de Agronomia Universidade Federal de Santa Maria, Santa Maria, 2011.

YANG, L. et al. Updating Conventional Soil Maps through Digital Soil Mapping. **Soil Science Society of America Journal**, Madison, v. 75, n. 3 p. 1044–1053, 2011.

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

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

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

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