

# SELEÇÃO DE DENSIDADE DE AMOSTRAGEM COM BASE EM DADOS DE ÁREAS JÁ MAPEADAS PARA TREINAMENTO DE MODELOS DE ÁRVORE DE DECISÃO NO MAPEAMENTO DIGITAL DE SOLOS

Tatiane Bagatini<sup>(1)\*</sup>, Elvio Giasson<sup>(2)</sup> e Rodrigo Teske<sup>(1)</sup>

<sup>(1)</sup> Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Departamento de Solos, Programa de Pós-graduação em Ciência do Solo, Porto Alegre, Rio Grande do Sul, Brasil.

<sup>(2)</sup> Universidade Federal do Rio Grande do Sul, Faculdade de Agronomia, Departamento de Solos, Porto Alegre, Rio Grande do Sul, Brasil.

\* Autor correspondente.

E-mail: tatibagatini@yahoo.com.br

## RESUMO

Para estudar técnicas de amostragem, úteis ao mapeamento digital de solos (MDS), objetivou-se avaliar o efeito da variação da densidade de pontos amostrais com base em dados de áreas já mapeadas por métodos tradicionais na acurácia dos modelos de árvores de decisão (AD) para a geração de mapas de solos por MDS. Em duas bacias hidrográficas no noroeste do Rio Grande do Sul, usou-se, como referência, antigos mapas convencionais de solos na escala 1:50.000. A partir do modelo digital de elevação do terreno e da rede hidrográfica, foram gerados mapas das variáveis preditoras: elevação, declividade, curvatura, comprimento de fluxo, acúmulo de fluxo, índice de umidade topográfica e distância euclidiana de rios. A escolha dos locais dos pontos amostrais foi aleatória e testaram-se densidades amostrais que variaram de 0,1 a 4 pontos/ha. O treinamento dos modelos foi realizado no *software Weka*, gerando-se modelos preditores usando diferentes tamanhos do nó final da AD para obter AD com tamanhos distintos. Quando não se controlou o tamanho das AD, o aumento da densidade de amostragem resultou no aumento da concordância com os mapas básicos de referências e no aumento do número de unidades de mapeamento preditas. Nas AD com tamanho controlado, o aumento da densidade de amostragem não influenciou a concordância com os mapas de referência e interferiu muito pouco no número de unidades de mapeamento preditas.

**Palavras-chave:** mapa de solos, SIG, modelo, predição.

Recebido para publicação em 16 de maio de 2014 e aprovado em 26 de fevereiro de 2015.

DOI: 10.1590/01000683rbc20140289

**ABSTRACT: SELECTION OF SAMPLING DENSITY BASED ON DATA FROM AREAS ALREADY MAPPED FOR TRAINING DECISION TREE MODELS IN DIGITAL SOIL MAPPING**

*In order to study sampling techniques useful for digital soil mapping (DSM), we evaluated the effect of changes in sampling density, based on data from areas already mapped by traditional methods, in regard to the accuracy of decision trees models for generating soil maps using DSM. In two watersheds in northwestern Rio Grande do Sul, Brazil, 1:50,000 scale conventional soils maps were used as reference maps. From the ASTER - GDEM Global Digital Elevation Model and the hydrographic network, maps of predictive variables were generated: elevation, slope, curvature, flow length, flow accumulation, topographic wetness index, and Euclidian distance of the streams. We used random sampling, and tested sampling densities that ranged from 0.1 to 4 points per hectare. Models were trained using Weka software, generating predictive models using different sizes of the final node to obtain decision trees of different sizes. The results indicate that when the size of the decision tree was not controlled, an increase in sampling density resulted in greater overall accuracy in accordance with the basic reference maps and an increase in the number of predicted soil mapping units. When the size of decision trees was controlled, an increase in sampling density did not affect the overall accuracy and had a very slight influence on the number of predicted mapping units.*

*Keywords: soil map, GIS, model, prediction.*

## INTRODUÇÃO

O mapeamento digital do solo (MDS) é definido como a criação e população de sistemas de informação pedológica espacial por meio do uso de métodos observacionais de campo e laboratório, acoplados a sistemas de inferência pedológica espaciais e não espaciais (Lagacherie e McBratney, 2006). A sua utilização com base em sistemas de informação geográfica (SIG), estatística e pedologia está aumentando nas últimas décadas em razão do aumento de fontes de dados numéricos, como aqueles fornecidos pelos modelos digitais de elevação da superfície do solo (MDE), combinado com o desenvolvimento de novas ferramentas de processamento de informação (McBratney et al., 2003). A predição de tipos de solos na paisagem a partir dessa técnica consiste na utilização de modelos matemáticos que conseguem descrever as relações entre os fatores de formação do solo e suas ocorrências na paisagem. Entre esses modelos de predição, os que vêm ganhando destaque são os algoritmos de aprendizagem de máquinas. Nesse sentido, Qi e Zhu (2003), comparando três desses algoritmos, concluíram que árvores de decisão (AD) possuem maior capacidade de predição de atributos ou tipos de solos. Contudo, apesar de estudos de Moran e Bui (2002), Grinand et al. (2008), Giasson et al. (2011), entre outros, indicarem que os modelos de AD são adequados para uso no MDS, alguns procedimentos metodológicos ainda precisam ser definidos. Entre eles está a densidade de pontos amostrais nos mapas de variáveis ambientais e nos mapas de referência, onde são coletadas informações sobre a distribuição dos solos e que são usados para o treinamento dos modelos preditores (McBratney et al., 2003). Os mapas de referência são mapas pedológicos convencionais

que poderiam servir como base para mapeamento digital de áreas adjacentes ainda não mapeadas, e a amostragem nos mapas pedológicos e das variáveis ambientais é feita coletando-se em cada ponto amostral as informações representadas em cada mapa.

A seleção da densidade de amostragem merece ser investigada porque o número de amostras usadas no treinamento dos modelos pode influenciar significativamente a capacidade de predição dos algoritmos e sua acurácia (Hjort e Marmion, 2008). Nesse sentido, Zhu (2000) sugeriu adotar um número mínimo de amostras de pelo menos 30 vezes o número de unidades de mapeamento de solos existentes ou a serem preditas. Grinand et al. (2008), usando mapa na escala 1:250.000, apontaram que a densidade amostral sobre o mapa de solos equivalente a um ponto a cada 10 ha de área no terreno foi suficiente para captar a variabilidade de tipos de solos. Da mesma forma, em estudos realizados em MDS no Brasil recentemente há grandes diferenças quanto à densidade de amostragem sobre mapas de referência que tem sido utilizada, existindo estudos que utilizaram desde amostragem no mapa equivalente a 0,1 (Figueiredo et al., 2008) até 11 amostras por hectare (Carvalho et al., 2009). Assim sendo, o objetivo deste trabalho foi determinar o efeito da densidade de amostragem de dados nos mapas pedológicos e de variáveis ambientais no treinamento de modelos preditores sobre a acurácia dos modelos de AD e sobre o número de unidades de mapeamento de solo (UMS) predito na geração de mapas digitais de tipos de solos em duas bacias hidrográficas no noroeste do Estado do Rio Grande do Sul, visando treinar esses modelos preditores com o propósito de gerar mapas digitais de solos para áreas adjacentes às áreas já mapeadas.

## MATERIAL E MÉTODOS

As áreas de estudo são as bacias hidrográficas do Rio Santo Cristo e do Rio Lageado Grande. Essas estão situadas no noroeste do Estado do Rio Grande do Sul e inseridas na Bacia Hidrográfica U30 e apresentam áreas de 837 km<sup>2</sup> e 531 km<sup>2</sup>, respectivamente (Figura 1). O clima da região é subtropical úmido, tipo Cfa de Köppen, com precipitação pluvial média anual de 1.778 mm e temperatura média anual de 18,5 °C. O material de origem da região é basalto da Formação Serra Geral.

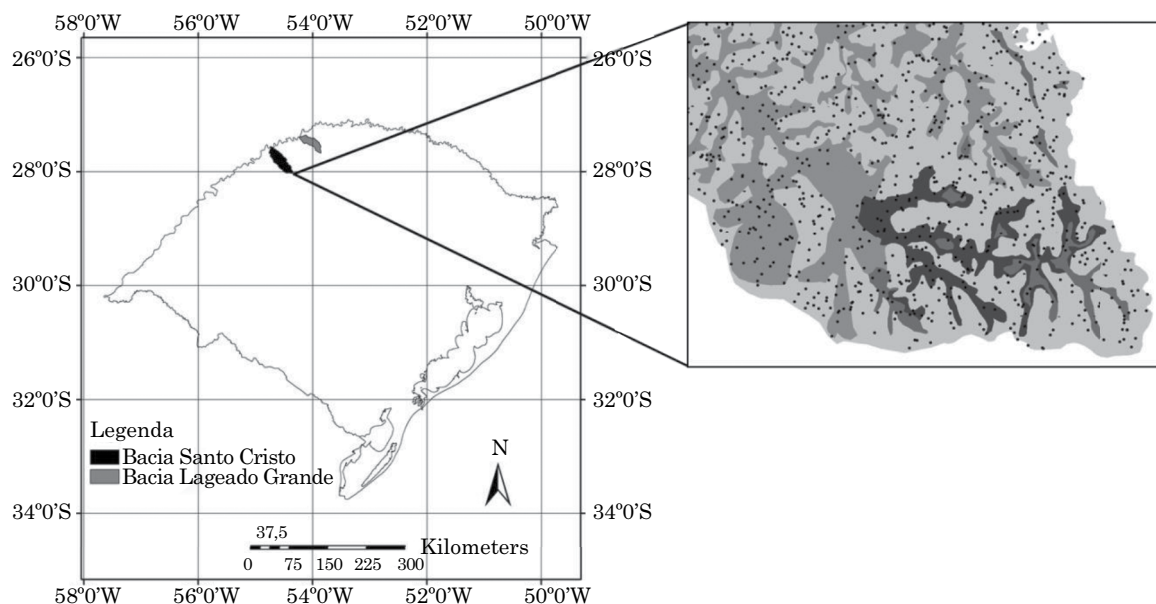
Os mapas de solos utilizados encontram-se na escala de 1:50.000 e fazem parte dos levantamentos pedológicos e da análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da bacia do Rio Santo Cristo (Kämpf et al., 2004a) e da bacia do Rio Lageado Grande (Kämpf et al., 2004b). Segundo esses mapas, na bacia do Rio Santo Cristo ocorrem 10 UMS e na do Rio Lageado Grande, 15 UMS. A composição das UMS e suas áreas de ocorrência nas bacias são apresentadas no quadro 1.

No programa ArcGis 9.3 (ESRI, 2008), a partir do MDE ASTER-GDEM com resolução de 30 m (Abrams et al., 1999), foram gerados mapas das seguintes variáveis preditoras dos solos na paisagem: declividade, direção do fluxo, acúmulo do fluxo, comprimento do fluxo, curvatura do declive e índice de umidade topográfica (Beven e Kirkby, 1979). Para geração da variável distância dos rios, foi utilizado o arquivo vetorial de hidrografia da base contínua do Rio Grande do Sul (Hasenack e Weber, 2010). Todos os mapas citados foram gerados com

resolução espacial de 30 m. Em razão de que cada uma das variáveis calculadas apresentam escalas de valores e unidades de medidas diferentes, as variáveis foram normalizadas, ficando com média zero e desvio-padrão um.

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

A amostragem propriamente dita foi realizada gerando-se o número de pontos aleatórios necessários para cada densidade amostral avaliada; nessas localizações espaciais, os dados sobre cada uma das variáveis foram coletados e tabulados utilizando-se a função *Sample* do ArcGis 9.3. Os dados tabulados consistiram de tabelas em que cada ponto amostral constituiu-se numa linha de tabela e cada coluna correspondia a uma variável amostrada. Essas tabelas foram exportadas para proceder



**Figura 1.** Localização das bacias hidrográficas dos Rios Santo Cristo e Lageado Grande no Estado do Rio Grande do Sul e esquema de distribuição dos pontos de amostragem, com densidade de 0,1 ponto por hectare.

**Quadro 1. Unidades de Mapeamento de Solos (UMS) das bacias do Rio Lageado Grande e do Rio Santo Cristo - classe de solos, área e representatividade de ocorrência das UMS**

Unidade de mapeamento	Área	
	km <sup>2</sup>	%
Bacia do Rio Lageado Grande		
Latossolo Vermelho - LV1	144,9	27,2
Associação Latossolo Vermelho e Neossolo Regolítico - LV2	31,2	5,9
Nitossolo Vermelho - NV1	2,4	0,4
Associação Nitossolo Vermelho e Neossolo Regolítico - NV2	9,0	1,7
Chernossolo Argilúvico - M1	3,6	0,7
Associação Chernossolo e Neossolo Regolítico - M2	3,9	0,7
Neossolo Regolítico - RR1	161,8	30,4
Associação Neossolo Regolítico e Neossolo Litólico, relevo forte ondulado - RR2	125,9	23,7
Associação Neossolo Regolítico e Latossolo Vermelho - RR3	28,1	5,3
Associação Neossolo Regolítico e Chernossolo Argilúvico - RR4	4,7	0,9
Associação Neossolo Regolítico e Cambissolo e Nitossolo Vermelho - RR5	2,7	0,5
Associação Neossolo Regolítico e afloramento rochoso - RR6	2,6	0,5
Associação Neossolo Litólico e Neossolo Regolítico - RL	2,2	0,4
Gleissolo Háplico e Gleissolo Melânico - G	8,7	1,6
Total	531,0	100
Bacia do Rio Santo Cristo		
Latossolo Vermelho - LV1	319,1	38,1
Associação Latossolo Vermelho e Neossolo Regolítico - LV2	66,3	7,9
Chernossolo Háplico - M	1,7	0,2
Neossolo Litólico - RL	0,01	0,01
Associação Neossolo Regolítico e Cambissolo Háplico - RR1	291,1	34,8
Complexo Neossolo Regolítico e Neossolo Litólico - RR2	69,0	8,2
Associação Neossolo Regolítico e Latossolo Vermelho - RR3	65,6	7,8
Associação Neossolo Regolítico e Neossolo Litólico - RR4	0,3	0,03
Associação Neossolos Regolíticos e Cambissolos Háplicos e Latossolo Vermelho - RR5	3,2	0,4
Gleissolo Háplico e Gleissolo Melânico - G	20,6	2,5
Total	837,0	100

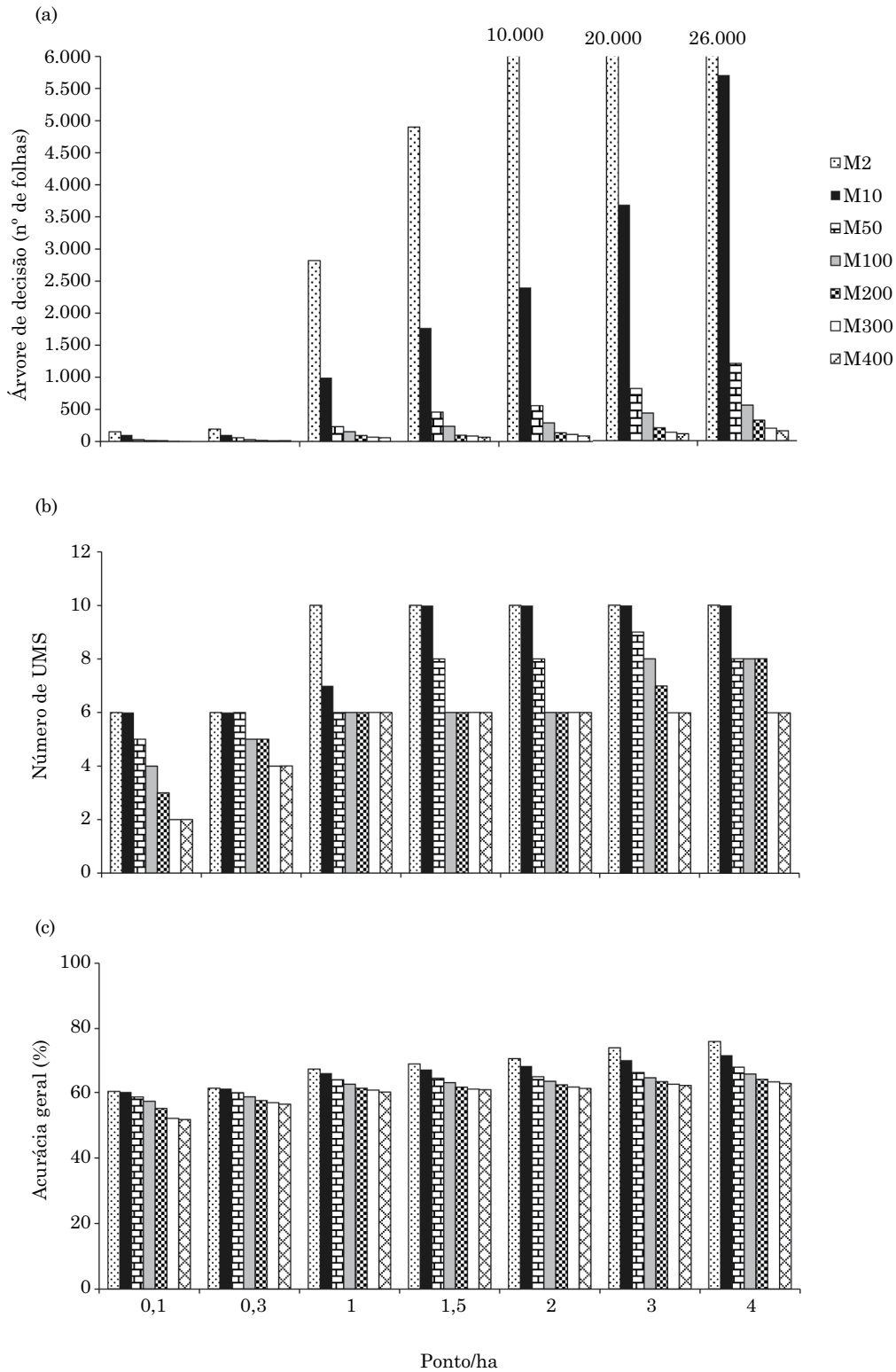
ao treinamento dos modelos de predição com o algoritmo *Simple Cart* no programa *Weka* 3.6.3 (Hall et al., 2009). A escolha do algoritmo *Simple Cart* deveu-se ao fato que, dos algoritmos testados (*J48* e *Simple Cart*) para as bacias em análise, foi o que resultou nos melhores valores de acurácia geral, concomitantemente com a geração de menores AD. Foram testados sete valores de número mínimo de elementos no nó final (M) nas AD no *Weka*, sendo esses: 2, 10, 50, 100, 200, 300 e 400. A variação do M intencionou avaliar o efeito da complexidade das AD na acurácia dos modelos preditores, dado que valores de M maiores geram AD menores, que são mais fáceis de implementar no ArcGis para a produção dos mapas quando esse processo é feito de forma manual. O teste utilizado para a seleção do modelo foi o *Cross Validation* (10 *folds*) e foi calculada a acurácia geral (AG) a partir da matriz de erros de Congalton (1991), que representa a concordância

entre o mapa de referência e a aplicação do modelo na geração de um mapa digital de solos.

## RESULTADOS E DISCUSSÃO

Os resultados referentes ao tamanho das AD, às acurácias geral (AG) e ao número de UMS preditas nas diferentes densidades de pontos amostrais nas bacias dos Rios Santo Cristo e Lageado Grande, respectivamente, são apresentados nas figuras 2a, 2b, 2c, 3a, 3b e 3c.

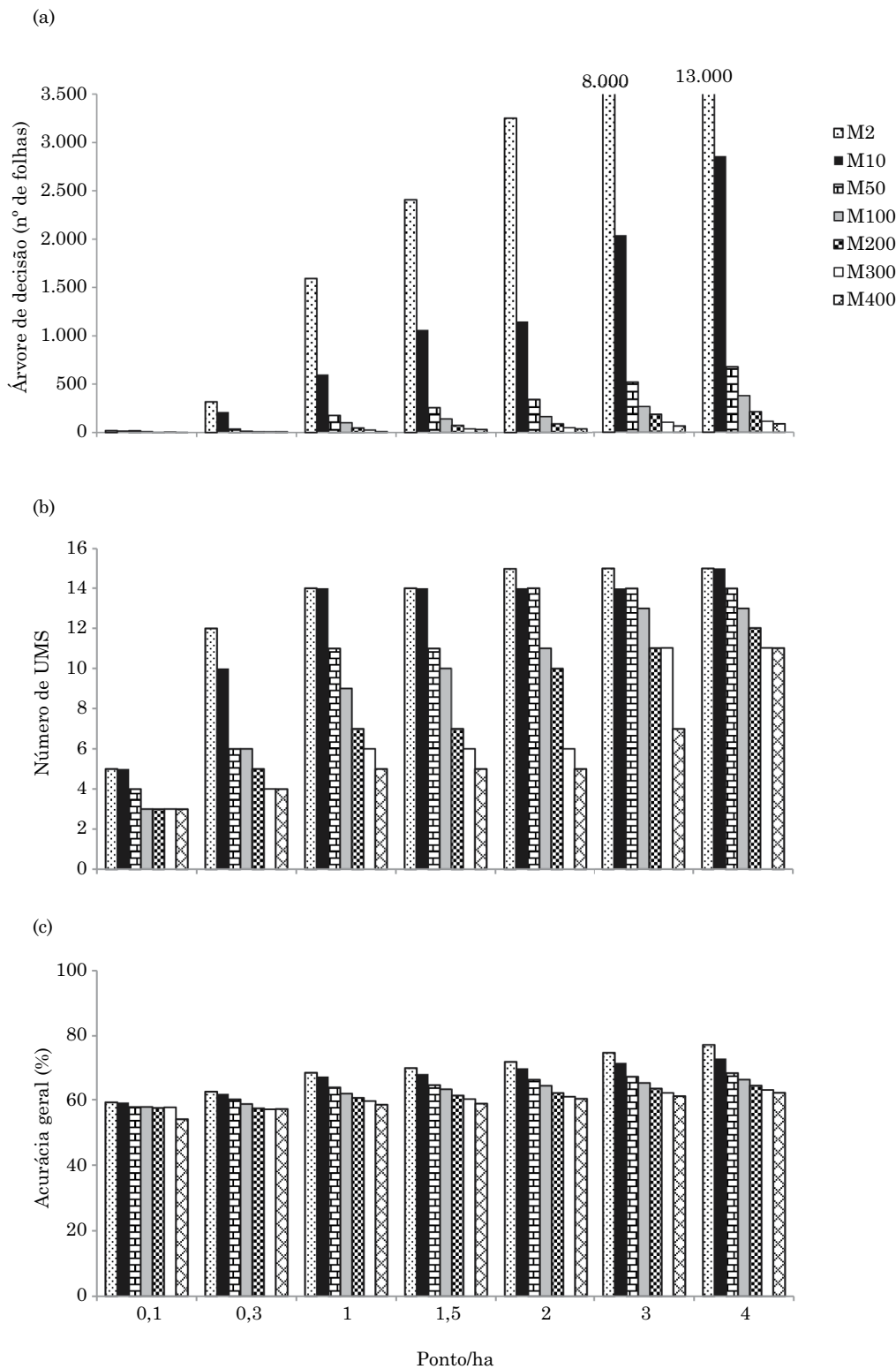
Em relação ao tamanho das AD da bacia do Rio Santo Cristo (Figura 2a), observou-se que com a elevação da densidade amostral houve aumento no tamanho das árvores com todos os valores de M testados, sendo geradas desde árvores com menos de 200 folhas, na densidade de 0,1 ponto/ha, até



**Figura 2. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos (UMS) previstas (b) e acurácia geral (c) nas diferentes densidades de amostragens na bacia do Rio Santo Cristo.**

árvores com mais de 25.000 folhas, na densidade de 4 pontos/ha. Quando analisaram-se os resultados dos tamanhos das AD na bacia do Rio Lageado

Grande (Figura 3a), observou-se que, de modo semelhante à bacia do Rio Santo Cristo, ocorreu aumento no número de folhas em razão da



**Figura 3. Tamanho das árvores de decisão (a), número de unidades de mapeamento de solos previstas (UMS) (b) e acurácia geral (c) nas diferentes densidades de amostragens na bacia do Rio Lageado Grande.**

elevação da densidade amostral, apresentando AD com menos de 100 folhas na densidade de 0,3 ponto/ha para mais de 14.000 folhas na

densidade de 4 pontos/ha. O aumento no tamanho das árvores com a elevação da densidade amostral também foi observado por Moran e Bui (2002) e

por Ten Caten et al. (2013), que destacaram que esses modelos de AD possuem grande capacidade preditiva, porém com grande complexidade.

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

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

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

De modo geral, observou-se que o aumento do tamanho das AD, principalmente com  $M = 2$ , em todas as densidades, não refletiu no aumento proporcional das acurácias em ambas as bacias. Entretanto, quando se verificaram os dados referentes às AD menos complexas com tamanhos

entre 150 e 200 folhas, na bacia do Rio Santo Cristo, não se observaram grandes diferenças nos valores de acurácia (60 a 63 %) e nem diferenças no número de UMS preditas (seis) entre as diferentes densidades de amostragem. Já na bacia do Rio Lageado Grande, notou-se que, igualmente, nas AD menos complexas, não houve aumento na acurácia geral com a elevação da densidade de amostragem, porém ocorreu aumento do número de UMS preditas. Na menor densidade de amostragem (0,1 ponto/ha), foram preditas somente cinco UMS (G, LV1, LV2, RR1 e RR2), enquanto nas maiores foram preditas todas as 15 UMS.

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

Portanto, analisando os dados das duas bacias, observou-se que a utilização da densidade de amostras de pelo menos 1 ponto/ha nos mapas antes efetuados por métodos tradicionais é o suficiente para que a amostragem bem represente as áreas. No entanto, se houver a necessidade de controlar o tamanho das AD, é indiferente utilizar qualquer uma das densidades de amostragem para obter valores de acurácia geral maiores que 60 %.

## CONCLUSÕES

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

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

Na região estudada, o uso de densidades de amostragem de menos de uma amostra por hectare demonstrou ser inadequado para a geração de mapas digitais de solos, principalmente quando nos mapas de referência ocorreu grande número de UMS.

## AGRADECIMENTOS

Ao CNPq, pela bolsa de doutorado da primeira autora e pela bolsa de Produtividade em Pesquisa do segundo autor; à CAPES, pela bolsa de doutorado do terceiro autor; e ao professor Alberto Vasconcellos Inda Junior, pela colaboração na revisão deste artigo.

## REFERÊNCIAS

- Abrams M, Hook S, Ramachandran B. ASTER user's handbook. Pasadena: JPL; 1999.
- Beven K, Kirkby N. A physically based variable contributing area model of basin hydrology. *Hydrol Sci Bull Sci Hydrol*. 1979;24:43-69.
- Carvalho CCN, Rocha WF, Ucha JM. Mapa digital de solos: Uma proposta metodológica usando inferência fuzzy. *R Bras Eng Agríc Amb*. 2009;13:46-55.
- Congalton RG. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens Environ*. 1991;37:35-46.
- Environmental Systems Research Institute - ESRI. ArcGIS Desktop. Redlands: 2008.
- Figueiredo SR, Giasson E, Tornquist CG, Nascimento PC. Uso de regressões logísticas múltiplas para mapeamento digital de solos no planalto médio do RS. *R Bras Ci Solo*. 2008;32:2779-85.
- Giasson E, Sarmento EC, Weber E, Flores CA, Hasenack H. Decision trees for digital soil mapping on subtropical basaltic steplands. *Sci Agríc*. 2011;68:167-74.
- Grinand C, Arrouays D, Laroche B, Martin M. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*. 2008;143:180-90.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update. 2009. *SIGKDD Explor*. 2009;11:10-8.
- Hasenack H, Weber E. Base cartográfica vetorial contínua do Rio Grande do Sul - escala 1:50.000. [DVD-ROM]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2010. (Geoprocessamento, 3).
- Hjort J, Marmion M. Effects of sample size on the accuracy of geomorphological models. *Geomorphology*. 2008;102:341-50.
- Kämpf N, Giasson E, Streck EV. Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da bacia do Rio Santo Cristo [relatório]. Porto Alegre: Secretaria do Meio Ambiente do Rio Grande do Sul; 2004a.
- Kämpf N, Giasson E, Streck EV. Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da microbacia do Rio Lageado Grande. [relatório final] Porto Alegre: Secretaria do Meio Ambiente do Rio Grande do Sul; Programa Nacional do Meio Ambiente II, 2004b.
- Lagacherie P, Mcbratney AB. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie P, Mcbratney AB, Voltz M, editors. *Digital soil mapping - an introductory perspective*. Amsterdam: Elsevier; 2006. p.3-22.
- Mcbratney AB, Mendonça-Santos MI, Minasny B. On digital soil mapping. *Geoderma*. 2003;117:3-52.
- Moran CJ, Bui E. Spatial data mining for enhanced soil map modelling. *Int J Geogr Inf Sci*. 2002;16:533-49.
- Pal M, Mather PM. A comparison of decision tree and backpropagation neural network classifiers for land use classification. In: *Proceedings of the 2<sup>nd</sup> International Geoscience and Remote Sensing Symposium*. Nottingham: IEEE International; 2002. p.503-505.
- Qi F, Zhu AX. Knowledge discovery from soil maps using inductive learning. *Int J Geogr Inf Sci*. 2003;17:771-95.
- Ten Caten A, Dalmolin RSD, Pedron FA, Ruiz LFC, Silva CA. An appropriate data set size for digital soil mapping in Erechim, Rio Grande do Sul, Brazil. *R Bras Ci Solo*. 2013;37:359-66.
- Zhu AX. Mapping soil landscape as spatial continua: The neural network approach. *Water Resour Res*. 2000;36:663-77.