

Expansão de mapas pedológicos para áreas fisiograficamente semelhantes por meio de mapeamento digital de solos

Tatiane Bagatini⁽¹⁾, Elvio Giasson⁽¹⁾ e Rodrigo Teske⁽¹⁾

⁽¹⁾Universidade Federal do Rio Grande do Sul, Departamento de Solos, Avenida Bento Gonçalves, nº 7.712, CEP 91540-000 Porto Alegre, RS, Brasil. E-mail: tatibagatini@yahoo.com.br, giasson@ufrgs.br, rodrigoteske.agr@gmail.com

Resumo – O objetivo deste trabalho foi realizar a expansão de mapas pedológicos pela extrapolação de mapas preexistentes para áreas fisiograficamente semelhantes. Foram utilizados mapas de solos, em escala 1:50.000, das bacias hidrográficas dos rios Santo Cristo e Arroio Portão, no Rio Grande do Sul, e a extrapolação foi feita com uso do algoritmo de árvores de decisão “simple cart”, treinado nas áreas previamente mapeadas. As bacias foram divididas em duas partes, uma para o treinamento e outra para a validação do modelo. A partir do modelo digital de elevação Aster-GDEM, foram gerados sete mapas de variáveis preditoras dos solos na paisagem. A amostragem de dados foi aleatória, com densidade de três pontos por hectare. O treinamento dos modelos foi realizado no programa Weka, e as acurácias foram calculadas a partir de matriz de erros. Para ambas as bacias, a acurácia geral do mapa de solos predito foi maior na área de treinamento do que na área de validação, a qual apresentou valores de 50 e 54%. Os mapas produzidos pelo modelo preditor apresentaram acentuada diferença na distribuição espacial das unidades de mapeamento, comparados com o mapa de solos original, indicio de que a técnica de mapeamento digital utilizada é pouco eficiente para extrapolar mapas de solos preexistentes para outras áreas fisiograficamente semelhantes.

Termos para indexação: acurácia geral, árvores de decisão, mineração de dados, pedometria, simple cart, sistema de informação geográfica.

Expanding pedological maps to physiographically similar areas with digital soil mapping

Abstract – The objective of this work was to expand pedological maps by extrapolating existing soil maps to physiographically similar areas. Soil maps were used at the scale of 1:50,000, for the watersheds of the rivers Santo Cristo and Arroio Portão, in the state of Rio Grande do Sul, Brazil, and the extrapolation was done using the “Simple Cart” decision tree algorithm, trained in the previously mapped areas. The watersheds were divided into two parts, one used for model training and the other for model validation. From the digital elevation model Aster-GDEM, seven maps of soil predicting variables in the landscape were generated. Sampling was random and performed with sampling density of three points per hectare. Model training was performed in the Weka software, and model accuracies were calculated using the error matrix. For both watersheds, the overall accuracy of the predicted soil map was higher in the training area than in the validation area, and showed values of 50 and 54%. The maps produced by the predictive model showed acute differences in the spatial distribution of mapping units, compared with the original soil map, indicating that the used digital mapping technique has low effectivity for the extrapolation of pre-existing soil maps to other physiographically similar areas.

Index terms: overall accuracy, decision trees, data mining, pedometrics, simple cart, geographic information system.

Introdução

Apesar de uma crescente necessidade de informações sobre levantamentos de solos em média e grande escalas, os custos, os recursos humanos e o tempo necessários para a sua realização tornam esses levantamentos de difícil execução. Nesse aspecto, o Mapeamento Digital de Solos (MDS) tem se tornado uma alternativa para geração desses mapas por meio

de métodos matemáticos computacionais associados a variáveis ambientais derivadas do Modelos Digital de Elevação (MDE), do uso da terra, do material de origem, entre outros, em sistemas de informação geográfica (McBratney et al., 2003; Hengl et al., 2007).

A exemplo dos levantamentos pedológicos convencionais, a maior parte dos estudos com MDS baseia-se nas relações solo-paisagem. Entretanto, enquanto nos levantamentos convencionais essas

relações são estabelecidas de forma qualitativa, em uma avaliação abrangente e integrada dos diversos elementos do meio natural, no mapeamento digital elas são avaliadas quantitativamente por meio de variáveis ambientais, como as geomorfométricas, derivadas do MDE. Nos modelos digitais de elevação, relações específicas entre variáveis geomorfométricas e classes de solos são estabelecidas com auxílio de modelos matemáticos que visam à predição de ocorrência de classes de solos, por meio do treinamento desses modelos com mapas de solos convencionais (Lagacherie & McBratney, 2007). Portanto, o uso desses modelos pode permitir a expansão dos mapas de solos por meio da extrapolação do mapeamento existente para áreas fisiograficamente semelhantes (Grinand et al., 2008, Höfig et al., 2014).

Apesar da existência de várias técnicas para realização do MDS, as árvores de decisão (AD) (Breiman et al., 1984) têm sido cada vez mais usadas para esse propósito (Grinand et al., 2008; Giasson et al., 2011; Caten et al., 2013; Höfig et al., 2014), principalmente em razão de serem de fácil entendimento, interpretação e discussão, já que seguem a lógica usualmente empregada em classificações sistemáticas (Xu et al., 2005). Além disso, elas são capazes de processar grandes volumes de dados, praticamente sem a necessidade de pré-processamento ou transformação. Essa grande capacidade de tratamentos de dados, sem interferência humana, faz com que as ADs sejam capazes de recuperar conhecimento pedológico utilizando-se mapas existentes, o que as torna uma ferramenta eficiente de mineração de dados (Henderson et al., 2005).

Como os modelos de AD são treinados com o estabelecimento de relações entre ocorrência de solos e variáveis ambientais, eles podem ser usados para extrapolar relações solo-paisagem de uma área previamente mapeada, para outras fisiograficamente semelhantes ainda não mapeadas. Apesar de essa extrapolação ter sido sugerida por Lagacherie et al. (1995), até o momento, muitos dos estudos realizados com MDS têm buscado avaliar a capacidade dos modelos de reproduzir os mapas de solos existentes, mas não a sua capacidade de extrapolação. Em relação a isso, Grinand et al. (2008) constataram que a capacidade preditiva do modelo por eles utilizado foi bastante baixa na área de validação do modelo preditor. No Brasil, Caten et al. (2011) constataram que

a acurácia do modelo também foi menor na área de validação, em comparação com a área de treinamento, e que os melhores resultados foram obtidos em áreas mais declivosas, com predominância de Neossolos. Todavia, Höfig et al. (2014), por meio de avaliação da verdade de campo, reportaram acurácia do mapa predito semelhante à acurácia do mapa convencional de solos, que foi de 75%, e concordância do mapa predito com o mapa convencional de 47%.

O objetivo deste trabalho foi realizar a expansão de mapas pedológicos pela extrapolação de mapas preexistentes para áreas fisiograficamente semelhantes.

Material e Métodos

Dois bacias hidrográficas, localizadas em regiões distintas do Rio Grande do Sul, foram estudadas de maneira independente. As bacias apresentam diferenças quanto a: área, relevo, material de origem, clima e, conseqüentemente, solos.

A primeira bacia hidrográfica é a do Rio Santo Cristo, localizada na região fisiográfica do Alto Uruguai, com área aproximada de 83.700 ha. O clima da região é subtropical úmido, tipo Cfa de Köppen, com chuvas regulares durante todo o ano, com precipitação pluvial média anual de 1.778 mm e temperatura média anual de 18,5°C. O substrato geológico regional é constituído por basaltos da Formação Serra Geral. O relevo da região apresenta-se como um planalto profundamente recortado pelos afluentes do Rio Uruguai; suave ondulado nas proximidades das nascentes dos rios Santo Cristo e Tuparendi; e mais acidentado em direção ao Rio Uruguai, com vales profundos e encostas íngremes. O mapa de solos utilizado como base (escala de 1:50.000) faz parte do levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da microbacia do Rio Santo Cristo (Kämpf et al., 2004).

A segunda bacia hidrográfica é a do Arroio Estância Velha/Portão, que abrange aproximadamente 22.160 ha. Nessa bacia, ocorrem as formações geológicas Botucatu, Rosário do Sul e Serra Geral. O clima da região também é do tipo Cfa, segundo Köppen, com precipitação pluvial média anual de 1.400 mm e temperatura média de 19,3°C. O mapa de solos utilizado (1:50.000) faz parte do levantamento semidetalhado de solos na bacia de Arroio Portão (Klamt et al., 1992).

As áreas de referência foram subdivididas em duas partes, uma utilizada para o treinamento e a outra para a validação do modelo preditor (Figura 1). Tal divisão foi realizada de modo que todas as UMs dos mapas estivessem presentes em ambas as partes. A simbologia usada em ambos os mapas de solos, e a descrição das classes taxonômicas que compõem as unidades de mapeamento (UM) e as áreas de ocorrência de cada UM estão apresentadas na Tabela 1.

Em ambiente de sistemas de informações geográficas (SIG), com o programa ArcGis 9.3 (Environmental Systems Research Institute, Redlands, CA, EUA), foram geradas variáveis predictoras da distribuição dos solos na paisagem. Os atributos do terreno usados como variáveis predictoras foram: elevação, declividade, acúmulo do fluxo, comprimento do fluxo, curvatura e índice de umidade topográfica (Beven & Kirkby, 1979). Esses atributos foram obtidos a partir do MDE Aster-GDEM (Abrams et al., 1999), com resolução espacial de 30 m. Para a geração da variável “distância dos rios”, utilizou-se o arquivo vetorial de hidrografia da base contínua do Rio Grande do Sul (Hasenack & Weber, 2010). Como as variáveis

calculadas apresentam escalas de valores e unidades de medidas distintas, procedeu-se à normalização estatística dessas variáveis com a ferramenta “Raster Calculator”, do ArcGis 9.3, para que todas as variáveis ficassem numa mesma ordem de grandeza, com média zero e desvio padrão 1.

A amostragem de dados para treinamento dos algoritmos foi realizada de maneira aleatória, com

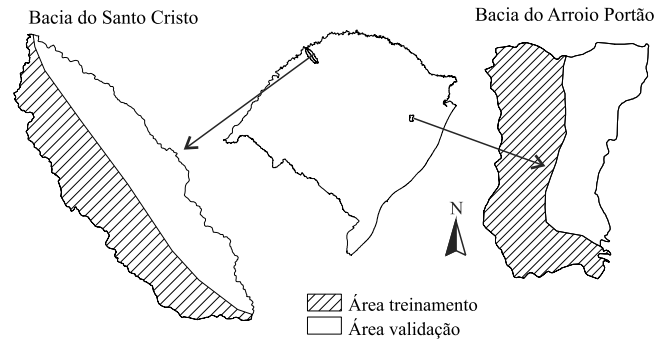


Figura 1. Localização das bacias do Santo Cristo e do Arroio Portão, com representação esquemática das áreas de treinamento e de validação.

Tabela 1. Composição taxonômica, áreas de ocorrência e legenda de identificação das Unidades de Mapeamento (UM) da bacia Santo Cristo e da bacia do Arroio Portão, no Rio Grande do Sul.

Unidade de mapeamento	Composição	Área (km ²)	Área (%)
Bacia do Santo Cristo			
LV1	Latossolo Vermelho	319,1	38,13
LV2	Associação de Latossolo Vermelho e Neossolo Regolítico	66,3	7,92
M	Chernossolo Háptico	1,7	0,2
RL	Neossolo Litólico	0,01	0,01
RR1	Associação de Neossolo Regolítico e Cambissolo Háptico	291,1	34,79
RR2	Complexo de Neossolo Regolítico e Neossolo Litólico	69	8,24
RR3	Associação de Neossolo Regolítico e Latossolo Vermelho	65,6	7,84
RR4	Associação de Neossolo Regolítico e Neossolo Litólico	0,3	0,03
RR5	Associação de Neossolo Regolítico, Cambissolo Háptico e Latossolo Vermelho	3,2	0,38
G	Associação de Gleissolo Háptico e Gleissolo Melânico	20,6	2,46
Total		837	100
Bacia do Arroio Portão			
CHa	Cambissolo Húmico	0,88	0,4
PVd	Argissolo Vermelho	47,56	21,4
PVa	Argissolo Vermelho-Amarelo	35,11	15,8
MXo/RLe	Associação de Chernossolo Háptico e Neossolo Litólico	13,72	6,0
SXd	Planossolo Háptico	1,21	0,6
Aluvial	Solos hidromórficos	0,61	0,3
CXf/RLe	Associação de Cambissolo Háptico e Neossolo Litólico	12,25	5,8
GMd	Gleissolo Melânico	17,68	8,4
PVa/GMe	Associação de Argissolo Vermelho e Gleissolo Melânico	92,56	41,3
Total		221,6	100

densidade de três pontos por hectare, em ambas as bacias. Para a Bacia do Santo Cristo, foram gerados 130.000 pontos amostrais, e, para a do Arroio Portão, 30.000 pontos. Nesses pontos, foram amostrados valores de todas as variáveis preditoras citadas acima. Os dados foram exportados para se proceder ao treinamento dos modelos com o algoritmo “Simple Cart”, no programa Weka 3.6.3 (Hall et al., 2009), tendo-se selecionado o modelo mais adequado pelos critérios das acurácias (Congalton, 1991) e do tamanho da árvore de decisão (AD). A aplicação dos modelos de AD foi feita no ambiente SIG, para predizer a distribuição espacial das UMs de solos nas áreas de treinamento e validação, tendo-se gerado o mapa digital de solos.

A acurácia dos modelos preditores foi avaliada nas áreas de treinamento e de validação. A concordância entre as áreas de ocorrência das UMs no mapa digital gerado – resultante da aplicação desses modelos – e nos mapas de solos originais foi avaliada com uso de matrizes de erros (Congalton, 1991). A concordância teve três denominações: AG (acurácia geral); AU (acurácia do usuário); e AM (acurácia do mapeador). Esta última expressa o quanto da área de uma UM do mapa convencional também foi mapeado da mesma forma no mapa gerado por MDS. A acurácia do usuário expressa a chance de uma UM mapeada no MDS, referente a um determinado pixel, coincidir com a UM do mapa convencional de solos. A acurácia geral expressa a proporção de classificações corretas em relação ao número total de pixels.

Resultados e Discussão

O mapa original de solos da bacia do Santo Cristo é composto por 10 UMs. As unidades LV1 e LV2 estão associadas às partes da bacia com relevo suave a ondulado, enquanto as demais estão relacionadas a relevo ondulado até fortemente ondulado (Tabela 2). Observa-se que o modelo preditivo foi capaz de reproduzir 63% do mapa original, na área de treinamento, e que a reprodutibilidade baixou para 50%, quando a extrapolação foi realizada para a área de validação. De modo geral, a AM foi maior na área de treinamento, em praticamente todas as UMs, em comparação com a área de validação. Entre as UMs preditas na área de treinamento, verifica-se que os maiores valores de AM estiveram associados às unidades LV1 e RR1, com 74,3 e 71%, respectivamente.

Isso significa que estas foram especializadas de forma muito semelhante ao que ocorre no mapa convencional de solos. A mesma tendência foi observada para a área de validação, porém, com valores menores, de 58% para LV1 e 66,1% para RR1.

As UMs RR4 e RR5 não apareceram na área de treinamento porque ocorrem somente na área de validação e, portanto, seus dados não foram inseridos para a construção do modelo. A unidade RL aparece com informação nula pelo fato de ocorrer em proporção muito inferior às demais, tendo sido classificada como RR1 e RR2. Os mesmos resultados foram observados para a RR4, na área de validação, onde essa UM também foi classificada como RR1 e RR2. Na área de validação, a RL não foi predita porque ela era encontrada somente na área de treinamento.

Como os erros de classificação foram inseridos nas UMs LV1 e RR1, ambas acabaram sendo superestimadas na área de treinamento, e passaram de 30% para 32,7% na primeira, e de 39,2% para 43,2% na segunda. Entretanto, na área de validação, os erros de classificação foram inseridos, principalmente, na RR1, o que aumentou sua área em 10 pontos percentuais (de 30,1% para 40,5%) e, com isso, a área da LV1 diminuiu de 46,8% para 35,8%.

Os maiores valores de AM associados a LV1 e RR1 ocorreram em virtude da maior representatividade dessas UMs no mapa convencional (38 e 35% da área total, respectivamente). Esse comportamento também foi observado em estudo realizado por Hengl et al. (2007) e Sarmiento (2010), em que as UMs com maior extensão no mapa original de solos predominaram no final das predições. Outro fator que pode ter influenciado, principalmente a UM RR1 (Associação de Neossolo e Cambissolos), foi o fato de que seis das sete variáveis utilizadas para a geração dos modelos eram derivadas do relevo, e este tem forte relação com a ocorrência de Neossolos, muito comuns em áreas com grande declividade.

Esses resultados também podem ser depreendidos da análise das Figuras 2 A e B, nas quais se constata que a maior dificuldade do modelo ocorreu, principalmente, na parte centro-norte. Nessa região, ocorre a maior presença de UMs, bem como maior complexidade de distribuição delas. Essa região apresenta relevo mais declivoso. É possível observar que a UM LV2 foi classificada, principalmente, como RR1. Contudo, esse erro de classificação pode estar

associado ao fato de que, em ambas as associações, os Neossolos Regolíticos estão presentes. Assim, pelo fato de essas UMs se encontrarem em posições da paisagem muito semelhantes, a separação realizada no mapa convencional de solos não foi identificada pelo modelo. Outra UM que foi erroneamente predita pelo modelo, especialmente na área de validação, foi a RR3, representada pela cor azul.

Para a bacia do Arroio Portão, a capacidade do modelo preditivo em reproduzir o mapa da área de treinamento apresentou uma reprodutibilidade de 68% (Tabela 3); porém, quando o modelo foi aplicado à área de validação, obteve-se a reprodutibilidade de 54%. De modo geral, a AM foi maior na área de treinamento em praticamente todas as UMs, quando comparada à área de validação. Na área de treinamento, verificou-se que os maiores valores de AM estão associados às UMs PVd, MXo/RLe e GMd, com 74,8, 75 e 77,2%, respectivamente. Já em relação à área de validação,

as UMs que foram mais bem preditas pelo modelo foram: PVa/GMe e MXo/RLe, com valores de 91,9 e 66,9%, respectivamente. As UMs CHa, SXd e Aluvial não foram preditas pelos modelos, possivelmente por causa da sua pequena representatividade no mapa de solos original, com apenas 0,4, 0,6 e 0,3% da área, respectivamente. As UMs CHa e Aluvial aparecem com valores nulos na área de treinamento, uma vez que estão situadas somente na área de validação.

Nessa bacia, as UMs MXo/RLe e CXf/RLe estão associadas a áreas de relevo mais declivoso, enquanto as demais estão associadas a relevo plano a suave ondulado. Assim, o modelo não conseguiu captar as diferenças para separação de UMs associadas a relevo menos declivoso, principalmente para a separação entre PVa e PVd, na área de validação, em que, na maioria dos casos, elas foram classificadas como PVa/GMe. Isso quer dizer que o modelo conseguiu prever a ocorrência de Argissolos ou os colocou

Tabela 2. Matriz de erros para acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG), referentes às áreas de treinamento e de validação da Bacia do Santo Cristo, no Rio Grande do Sul.

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

em uma associação que continha a classe. Resultados semelhantes foram observados por Figueiredo (2006), porém, com uso de regressões logísticas como ferramentas de predição. Essas diferenças também podem ser visualizadas nas Figuras 3 A e B, nas quais se constatou maior dificuldade de resposta do modelo principalmente na região centro-sul. Percebe-se que o modelo não conseguiu separar a UMPV_a, e a classificou como PVA/GMe, que são UMs predominantes.

Nas duas bacias, observou-se que a AG foi maior na área de treinamento, com área de validação 13% menor na bacia do Santo Cristo, e 14% menor na bacia do Arroio Portão. Grinand et al. (2008) observaram diferença de acurácia entre as áreas de 40%. Caten et al. (2011) também observaram diferença de 22,5% entre as áreas, com uso de regressão logística múltipla. Os autores argumentaram que a baixa acurácia na área de validação poderia resultar da complexidade inerente à distribuição espacial do solo na paisagem. Lemercier

et al. (2012), ao avaliarem a predição da drenagem natural do solo e do material de origem, observaram diferenças entre as áreas de treinamento e de validação de 18 e 24%, respectivamente. Entretanto, apesar de as diferenças na AG entre as áreas de treinamento e de validação no presente trabalho serem menores do que os valores relatados por esses autores, os valores de AG encontrados nas áreas de extrapolação, de 50 e 54%, foram superiores aos encontrados por Grinand et al. (2008) e Caten et al. (2011), mas inferiores aos encontrados por Höfig et al. (2014), os quais obtiveram AG de 75%. Esses resultados indicam que essa técnica ainda necessita de aprimoramento.

No presente trabalho, as diferenças de AG entre as áreas de treinamento e de validação podem ter ocorrido em virtude da complexidade da distribuição do solo na paisagem, conforme observado por Caten et al. (2011).

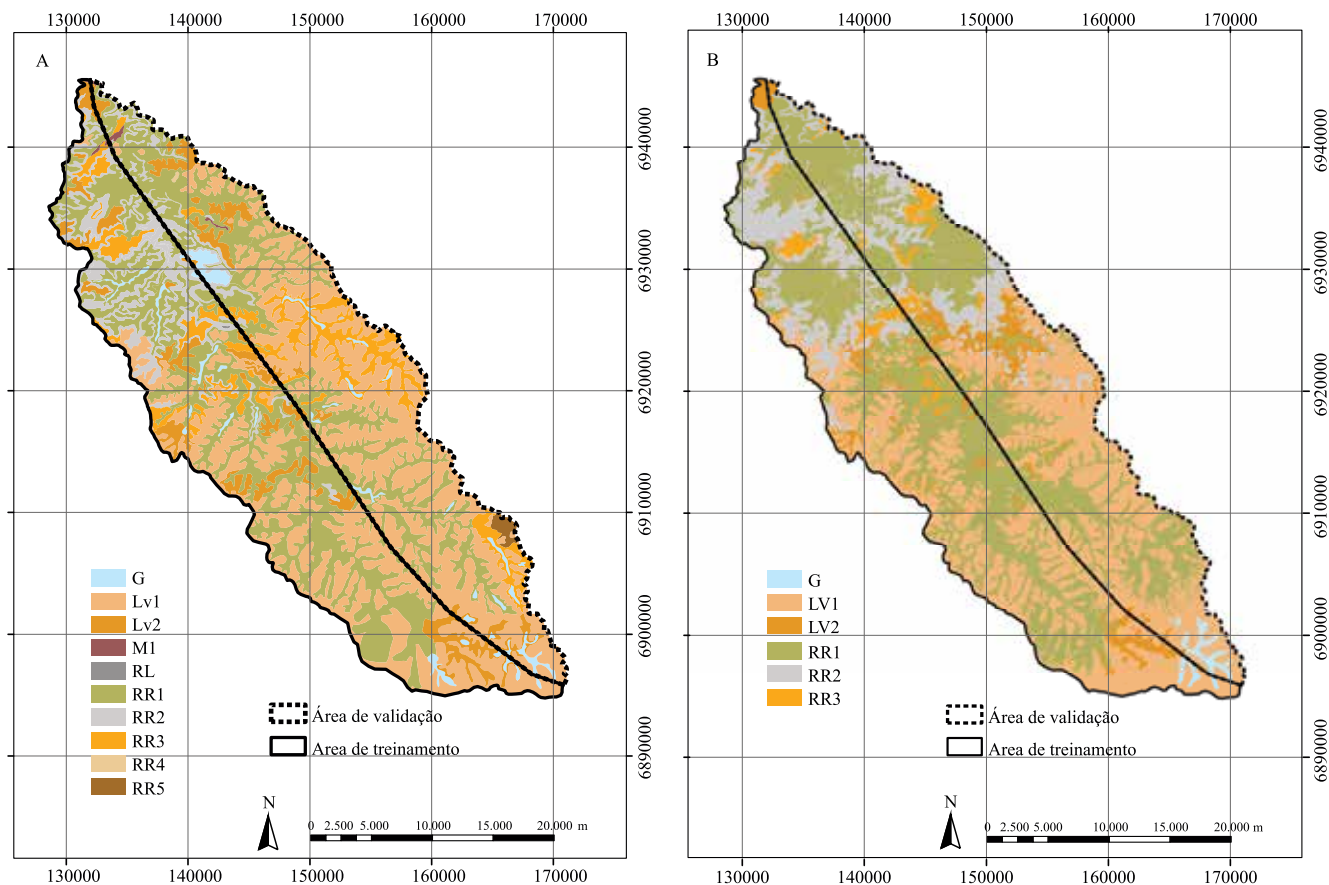
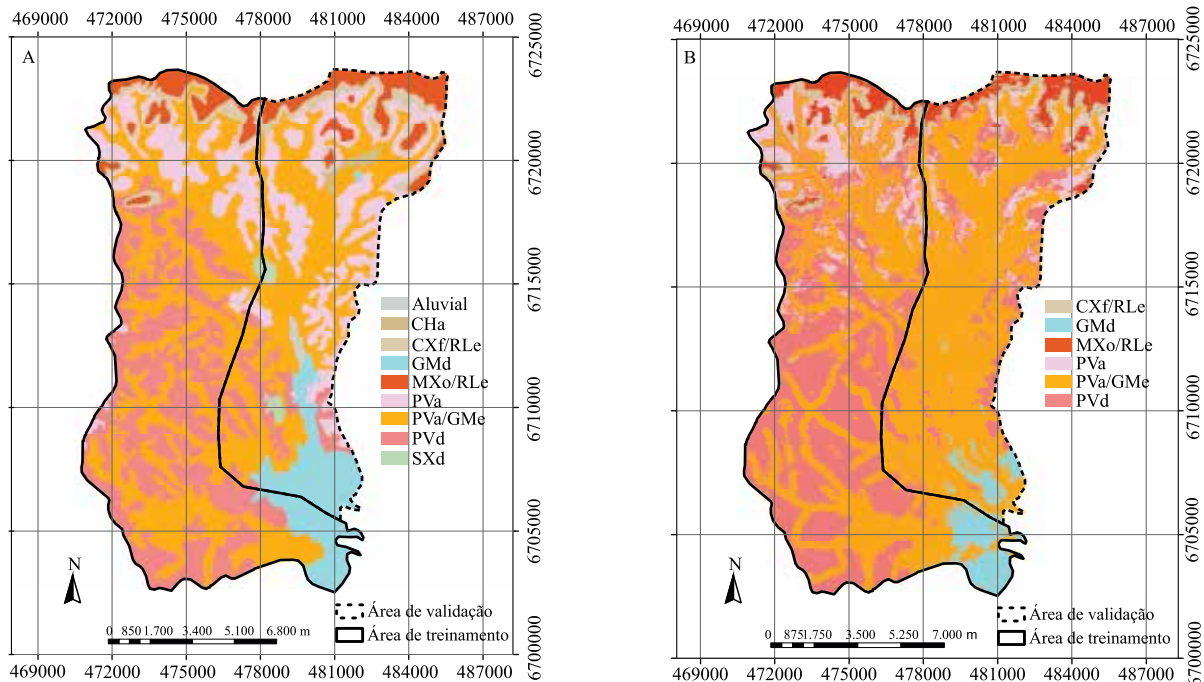


Figura 2. Mapa da bacia Santo Cristo com a distribuição original das unidades de mapeamento de solos (A) e o mapa digital de solos predito (B).

Tabela 3. Matriz de erros para acurácia do usuário (AU), acurácia do mapeador (AM), acurácia geral (AG), referentes às áreas de treinamento e de validação da Bacia do Arroio Portão, no Rio Grande do Sul.

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

**Figura 3.** Mapa da bacia Arroio Portão com a distribuição original das unidades de mapeamento de solos (A) e o mapa digital de solos predito (B).

Conclusões

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

2. Em ambas as bacias estudadas, os modelos superestimam as unidades de mapeamento com maior representatividade, e subestimam as de menor representatividade, tanto na área de treinamento como na de validação.

3. Nas condições deste estudo, a utilização de árvores de decisão para o mapeamento digital de solos a partir de uma área de treinamento mostra baixa eficiência para a extrapolação de mapas de solos preexistentes.

Referências

- ABRAMS, M.; HOOK, S.; RAMACHANDRAN, B. **ASTER users handbook**. Pasadena: JPL, 1999. 93p.
- BEVEN, K.J.; KIRKBY, M.J. A physically based, variable contributing area model of basin hydrology. **Hydrological Sciences Bulletin**, v.24, p.43-69, 1979. DOI: 10.1080/02626667909491834.
- BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A.; STONE, C.J. **Classification and regression trees**. Belmont: Wadsworth International Group, 1984.
- CATEN, A. ten; DALMOLIN, R.S.D.; PEDRON, F. de A.; RUIZ, L.F.C.; SILVA, C.A. da. An appropriate data set size for digital soil mapping in Erechim, Rio Grande do Sul, Brazil. **Revista Brasileira de Ciência do Solo**, v.37, p.359-366, 2013. DOI: 10.1590/S0100-06832013000200007.
- CATEN, A. ten; DALMOLIN, R.S.D.; PEDRON, F. de A.; SANTOS, M. de L.M. Componentes principais como preditores no mapeamento digital de classes de solos. **Ciência Rural**, v.41, p.1170-1176, 2011. DOI: 10.1590/S0103-84782011000700011.
- CONGALTON, R.G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991. DOI: 10.1016/0034-4257(91)90048-B.
- FIGUEIREDO, S.R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. 2006. 92p. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Porto Alegre.
- GIASSON, E.; SARMENTO, E.C.; WEBER, E.; FLORES, C.A.; HASENACK, H. Decision trees for digital soil mapping on subtropical basaltic steeplands. **Scientia Agricola**, v.68, p.167-174, 2011. DOI: 10.1590/S0103-90162011000200006.
- GRINAND, C.; ARROUAYS, D.; LAROCHE, B.; MARTIN, M.P. Extrapolating regional soil landscape from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. **Geoderma**, v.143, p.180-190, 2008. DOI: 10.1016/j.geoderma.2007.11.004.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I.H. The WEKA Data Mining Software: an update. **SIGKDD Explorations**, v.11, p.10-18, 2009. DOI: 10.1145/1656274.1656278.
- HASENACK, H.; WEBER, E. (Org). **Base cartográfica vetorial contínua do Rio Grande do Sul - escala 1:50.000**. Porto Alegre: UFRGS, Instituto de Biociências, Centro de Ecologia. 2010. 1 DVD. (Série Geoprocessamento, 3).
- HENDERSON, B.L.; BUI, E.N.; MORAN, C.J.; SIMON, D.A.P. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v.124, p.383-398, 2005. DOI: 10.1016/j.geoderma.2004.06.007.
- HENGL, T.; TOOMANIAN, N.; REUTER, H.I.; MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. **Geoderma**, v.140, p.417-427, 2007. DOI: 10.1016/j.geoderma.2007.04.022.
- 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**, v.49, p.958-966, 2014. DOI: 10.1590/S0100-204X2014001200006.
- KÄMPF, N.; GIASSON, E.; STRECK, E.V. **Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da bacia do Rio Santo Cristo**. Porto Alegre: Secretaria do Meio Ambiente do Rio Grande do Sul, 2004. 27p. Relatório final.
- KLAMT, E.; SCHNEIDER, P.; AZEVEDO, A.C.; SABADIN, E.M.; MOBUS, G.; SOARES, J.M.D.; CARPENEDO, V. **Solos da Bacia do Arroio Portão**: características, distribuição e aptidão de uso. Porto Alegre: Departamento de Solos da UFRGS, 1992. p.20.
- 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**, v.65, p.283-301, 1995. DOI: 10.1016/0016-7061(94)00040-H.
- 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.3-24. (Developments in Soil Science, 31). DOI: 10.1016/S0166-2481(06)31001-X.
- LEMERCIER, B.; LACOSTE, M.; LOUM, M.; WALTER, C. Extrapolation at regional scale of local soil knowledge using boosted classification trees: a two-step approach. **Geoderma**, v.171-172, p.75-84, 2012. DOI: 10.1016/j.geoderma.2011.03.010.
- MCBRATNEY, A.B.; MENDONÇA SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, p.3-52, 2003. DOI: 10.1016/S0016-7061(03)00223-4.
- SARMENTO, E.C. **Comparação entre quatro algoritmos de aprendizagem de máquina no mapeamento digital de solos no Vale dos Vinhedos, RS, Brasil**. 2010. 124p. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Porto Alegre.

XU, M.; WATANACHATURAPORN, P.; VARSHNEY, P.K.; ARORA, M.K. Decision tree regression for soft classification of remote sensing data. **Remote Sensing of Environment**, v.97, p.322-336, 2005. DOI: 10.1016/j.rse.2005.05.008.

Recebido em 24 de agosto de 2015 e aprovado em 23 de maio de 2016