



Mapeamento digital de solos com base em árvores de decisão em sub-bacia hidrográfica de Minas Gerais⁽¹⁾.

Adélia Aziz Alexandre Pozza⁽²⁾; Marcelo Henrique Procópio Pelegrino⁽³⁾; Sérgio Henrique Godinho Silva⁽⁴⁾; Michele Duarte de Menezes⁽⁵⁾; João José Marques⁽⁵⁾; Nilton Curi⁽⁶⁾

⁽¹⁾ Trabalho executado com recursos do PIBIC/UFLA, CNPq, CAPES E FAPEMIG.

⁽²⁾ Professora Adjunta; Depto. de Ciência do Solo (DCS) - Universidade Federal de Lavras (UFLA); Lavras, MG; adelia.pozza@dcs.ufla.br. ⁽³⁾ Estudante de Eng. Ambiental e Sanitária; UFLA; ⁽⁴⁾ Doutorando em Ciência do Solo; DSC - UFLA; ⁽⁵⁾ Professor(a) Adjunto(a); DCS-UFLA; ⁽⁶⁾ Professor Titular; DCS-UFLA.

RESUMO: Dados legados de solos são importante de informações e podem servir de base para novos mapeamentos de solos, a partir de técnicas de mineração de dados, como árvores de decisão. Este trabalho objetivou gerar árvores de decisão, com diferentes formas de extração de dados para sua construção, para mapeamento digital de solos de uma sub-bacia hidrográfica de Minas Gerais. A partir de um modelo digital de elevação (MDE), foram gerados atributos de terreno dos quais extraíram-se informações de cada pixel por classe de solo de um mapa pedológico na escala de 1:10.000. Foram extraídas informações: 1) do polígono como um todo e das partes mais centrais dos polígonos (20 e 40 m da borda) e 2) fazendo-se buffers dos pontos amostrados com 4 raios (25, 50, 75 e 100 m), como base para a geração das árvores de decisão. Mapas de solos foram gerados para cada árvore de decisão. A validação deles foi realizada com 20 pontos visitados em campo. As árvores geradas pelos buffers dos pontos tiveram menor acurácia por esses terem sido coletados em locais de transição entre polígonos, incluindo regiões de maior incerteza no mapeamento convencional. A árvore gerada pela extração de dados excluindo-se 20 m da borda da unidade obteve melhor resultado na predição de solos, com Índice Kappa 0,3525 e Índice Global de 55%. A extração de dados mais centrais nos polígonos para construir a árvore de decisão, mostrou-se mais adequada do que a extração de dados a partir de pontos com diferentes raios.

Termos de indexação: atributo de terreno, legado de solos, geoprocessamento.

INTRODUÇÃO

No Brasil, os recursos destinados ao levantamento de solos são escassos, onde mapas mais detalhados são comuns em pequenas áreas e geralmente atendem projetos específicos

(Mendonça Santos & Santos, 2007). Na falta de recursos para novos mapas, uma alternativa seria o uso de informações pré-existentes (Bui & Moran, 2001), o chamado legado de solos.

Os dados legados de solos, que incluem mapas com legendas, perfis modais ou pontos amostrados, podem ser muito úteis, onde informações a respeito do modelo mental do pedólogo podem ser recuperadas (Bui, 2004). No Brasil, a maior parte do legado de informações de solos encontra-se em mapas convencionais analógicos, onde o modelo mental do pedólogo não é explícito.

O legado de solos pode ser explorado pela árvore de decisão, um modelo preditivo empregado no mapeamento digital de solos. A extração das relações solo-paisagem, a partir de, por exemplo, atributos do terreno em formato raster, torna tais informações passíveis de predizerem solos, além de tornarem o processo mais documentado e explícito (Giasson et al., 2011). Neste processo, as árvores de decisão têm como principais vantagens a sua fácil interpretação e discussão, relativa simplicidade em modelar relações não aditivas e não lineares, capacidade de lidar com dados categóricos e numéricos (Greve et al., 2010), sendo o processo de sua aquisição relativamente rápido. As árvores de decisão auxiliam ainda na escolha dos atributos do terreno que se configuram como melhores preditores de classes de solos, auxiliando na compreensão da ocorrência dos solos na paisagem.

Este trabalho objetivou testar árvores de decisão, com diferentes quantidades de dados de entrada para o mapeamento de classes de solos da sub-bacia hidrográfica do Ribeirão Vista Bela (MG).

MATERIAL E MÉTODOS

A sub-bacia hidrográfica do Ribeirão Vista Bela localiza-se à margem direita do Rio Aiuruoca, no município de Andrelândia. Situa-se entre as coordenadas geográficas UTM 559.895 m e 7.597.109 m, e 561.563 m e 7.598.777 m, fuso 23K *datum* Córrego Alegre. Possui clima Cwa, segundo a classificação de Köppen, com inverno seco e

verão chuvoso. Sua altitude varia entre 960 e 1080 metros e nela ocorrem Latossolos Vermelho-Amarelos (34,4%), Latossolos Vermelhos (12%), Latossolos Amarelos (0,6%), Cambissolos Háplicos (35%), Neossolos Flúvicos (15,5%), Neossolos Litólicos (2,3%) e tem como material de origem gnaisses e biotita-xistos (Menezes, 2007).

Modelo digital de elevação (MDE) com resolução de 10 m foi gerado a partir de curvas de nível na escala de 1:50.000 (IBGE), no ArcGIS 10 com a ferramenta *Topo To Raster*. A partir do MDE, foram gerados quinze atributos de terreno (ATs): aspecto, declividade, distância vertical para a rede de drenagem, índice topográfico de umidade, índice de umidade SAGA, multiresolution index of valley bottom flatness, elevation, multiresolution index of ridge top flatness, plan e profile curvatures, hillshade, slope height e valley depth. Para tanto, foi empregado o SAGA GIS versão 2.1.4 (Böhner et al., 2006).

Os dados usados como legado foram obtidos a partir de levantamento pedológico convencional em escala 1:10.000 (Menezes, 2007). Tanto os polígonos das unidades de mapeamento como os pontos amostrais foram usados para extrair valores de pixel dos quinze ATs. Os procedimentos de extração de informações para construção das árvores de decisão foram os seguintes:

- 1) Polígonos das unidades de mapeamento: a) o polígono inteiro, b) interior do polígono com distância de 20 m em relação a borda do polígono; c) interior do polígono com distância de 40 m em relação a borda do polígono.
- 2) Pontos observados em campo: as informações foram obtidas a partir de um raio ou *buffer* no entorno dos pontos, variando a) raio de 25 m, b) raio de 50 m; c) 75 m; d) 100 m.

As árvores de decisão foram geradas através do *software* R (R DEVELOPMENT CORE TEAM, 2015), com o pacote *rpart* (Therneau et al., 2015).

A partir das informações das árvores, foram gerados os mapas de solos com a ferramenta *Raster Calculator* do *software* ArcGIS 10.1 (ESRI), totalizando sete mapas de solos. A acurácia dos mapas gerados foi medida através dos Índices Kappa e Global, calculados utilizando-se pontos de validação de campo para conferir as classes de solos em 20 locais.

RESULTADOS E DISCUSSÃO

Na **tabela 1** é possível notar que os diferentes métodos de extração geraram mapas de solos com acurácias contrastantes. Os mapas baseados nos diferentes raios a partir dos pontos apresentaram pior desempenho em relação aos que obtiveram

dados a partir dos polígonos. Isso se deve provavelmente ao método de amostragem na área de estudo para o mapeamento convencional cujo método amostral incluiu locais para checagem de limites entre as unidades de mapeamento, para que os polígonos fossem então delineados. Deste modo, os pontos utilizados para os buffers estão, em sua maioria, localizados em regiões de transição de classes de solos, o que pode ter prejudicado a modelagem a partir desta forma de extração de informações. Além disso, a contribuição das variáveis preditoras nestes locais para os modelos na árvore de decisão podem ser duvidosos (Ten Caten et al., 2012), o que também pode ter contribuído para a baixa acurácia dos pontos como base para extração da informação.

O polígono que desconsiderava 20 m da borda, que apresentou maior acurácia, representa o conceito mais central da unidade de mapeamento, distante das áreas das bordas, mas com uma quantidade maior de pixels para a construção da árvore, quanto comparado com a de 40 m (reduzido tamanho do buffer). Os limites (bordas) entre os polígonos são os locais que apresentam maiores incertezas no mapa de solos (Legros, 2006).

Na árvore de decisão gerada pelo polígono de 20 m (**Figura 1**), percebe-se que a altitude foi uma das principais variáveis preditoras. Esta árvore foi capaz de predizer todas as classes de solos encontradas no levantamento convencional de solos, conforme **figura 2**. No entanto, para a maioria das árvores geradas, a classe de solo Neossolo Litólico não foi predita, reduzindo a acurácia dos mapas gerados. Esta unidade de mapeamento apresenta reduzida área, conseqüentemente, os polígonos gerados para extração de informação são também reduzidos, o que pode contribuir para a não separação de tais áreas pela árvore de decisão.

O índice kappa de 0,3525 e acurácia global de 55% para o método que mostrou maior acurácia apontam que esta ferramenta tem potencial para mapeamento em escalas mais generalizadas. No entanto, deve-se ressaltar que este estudo partiu de um mapa em escala detalhada, cujas unidades de mapeamento são do tipo simples (predomínio de apenas uma unidade taxonômica), o que dificulta ainda mais a modelagem pelas árvores de decisão.

CONCLUSÕES

A extração de dados mais centrais nos polígonos para construir a árvore de decisão, mostrou-se mais adequada do que a extração de dados a partir de pontos com diferentes raios.

As árvores de decisão, apesar de terem se mostrado de fácil uso e rápida resposta para

individualização de cada classe de solo, não foram tão adequadas para o mapeamento em escala mais detalhada, conforme apontam os índices de validação. No entanto, todas as classes de solos foram contempladas pela árvore, mesmo aquelas unidade de mapeamento de menor expressão geográfica.

AGRADECIMENTOS

Os autores agradecem ao CNPq, a CAPES e FAPEMIG por fornecerem apoio financeiro para a realização deste trabalho.

REFERÊNCIAS

BÖHNER, J.; MCCLOY, K.R.; STROBL, J. SAGA - analysis and modeling applications. *Gottinger Geographische Abhandlungen*. 2006. v.115, 130p.

BUI, E.N. Soil survey as a knowledge system. *Geoderma* 120 (2004) 17–26.

BUI, E.N. & MORAN, C.J. Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. *Geoderma*, 103:79-94, 2001.

GIASSON, E. et al. Decision trees for digital soil mapping on subtropical basaltic steeplands. *Scientia Agricola*, 68:167-174, 2011.

GREVE, M.H. et al. Comparing Decision Tree Modeling and Indicator Kriging for Mapping the Extent of Organic Soils in Denmark. In: Boettinger et al. eds. *Digital Soil Mapping – Bridging research, environmental application, and operation*. Amsterdam:Springer, 2010, p.267-280.

LEGROS, J.P. *Mapping of the Soil*. Enfield: Science Publishers, 2006, 411p.

MENDONÇA-SANTOS, M.L.; MCBRATNEY A.B.; MINASNY B. 2007. Soil prediction with spatially decomposed environmental factors. In: LAGACHERIE, P.; MCBRATNEY, A.B.; VOLTZ, M. eds. *Digital soil mapping: an introductory perspective*, Melbourne:Elsevier Science, 2007, 31:269-278.

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. eds. *Digital soil mapping: an introductory perspective*, Melbourne:Elsevier Science, 2007. p.39-54

MENEZES, M.D. Levantamento de solos em sub- bacia hidrográfica como apoio para avaliação do uso da terra e da potencialidade de recarga de aquíferos. *Dissertação de Mestrado - Universidade Federal de Lavras, Lavras, 2007, 107p.*

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

TEN CATEN A.; DALMOLIN, R.S.D.; RUIZ, L.F.C. Digital soil mapping: strategy for data pre-processing. *Revista Brasileira de Ciência do Solo*, 36:1083-1091.

THERNEAU, T.; ATKINSON, B.; RIPLEY, B. *rpart: Recursive Partitioning and Regression Trees*. 2015. Acesso em 21/03/2015. Disponível em <http://cran.r-project.org/web/packages/rpart/rpart.pdf>.

Tabela 1 – Coeficiente kappa e acurácia global dos mapas gerados por árvore de decisão, a partir de diferentes métodos de extração de dados.

Método de extração dos dados	Polígono inteiro	Polígono 20 m	Polígono 40m	Pontos com raio de 25m	Pontos com raio de 50m	Pontos com raio de 75m	Pontos com raio de 100 m
Kappa	0,1971	0,3525	0,2517	0,0635	0,1195	-0,026	0,079
Global	45%	55%	45%	30%	30%	20%	30%

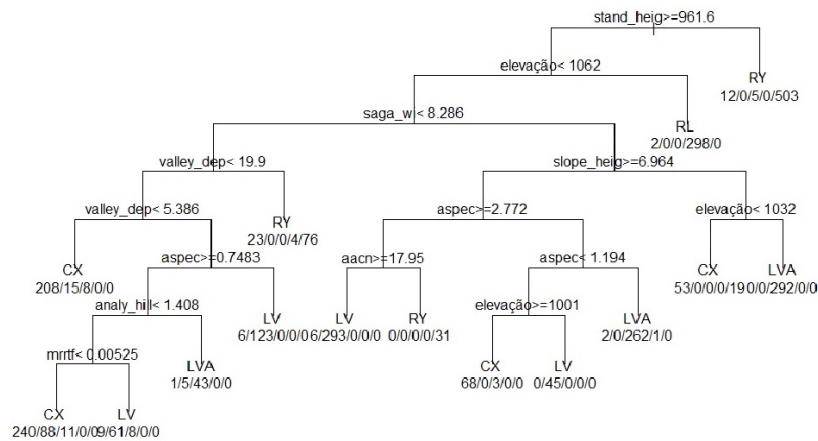


Figura 1 - Árvore de decisão gerada a partir de 20 m das bordas dos polígonos.

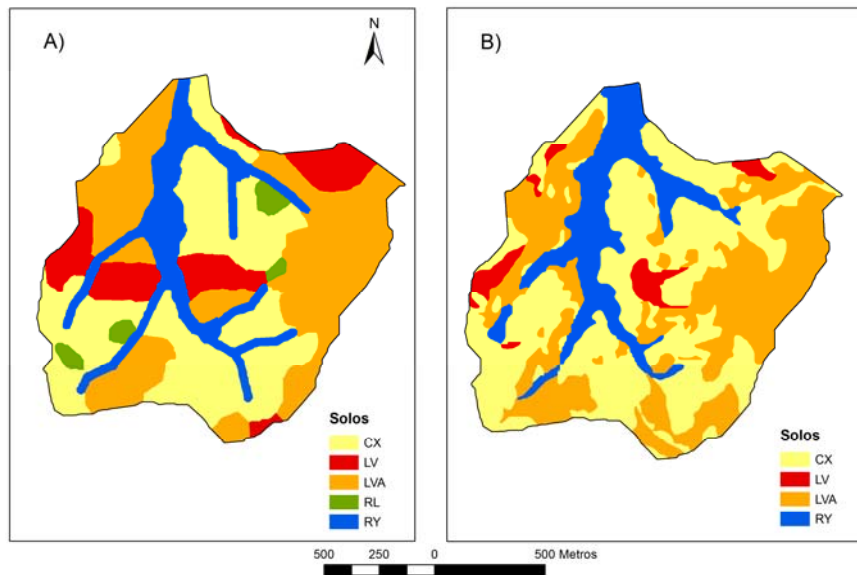


Figura 2 – Mapa de Solos da Sub-bacia do Ribeirão Vista Bela original (A) e obtido por árvore de decisão (B).