

Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil

Evaluation of five algorithms of decision trees and three digital elevation models for digital soil mapping at semidetail level at the Lageado Grande Watershed, RS, Brazil

**Elvio Giasson^{I*} Alfred Eduard Hartemink^{II} Carlos Gustavo Tornquist^I Rodrigo Teske^I
Tatiane Bagatini^I**

RESUMO

O mapeamento digital de solos (MDS) tem como base a geração de sistemas de informações que permitem estabelecer relações matemáticas entre variáveis ambientais e solos e, dessa forma, prever a distribuição espacial das classes ou propriedades dos solos. Dentre as abordagens mais utilizadas, as árvores de decisão têm se destacado por apresentar bons resultados no MDS. Por outro lado, dada a disponibilidade de novas fontes de informação sobre a elevação, torna-se necessário o teste e avaliação de modelos digitais de elevação (MDE) quanto ao seu uso para o MDS. Este estudo testa cinco algoritmos de árvores de decisão (Simple Chart, Random Tree, REP Tree, BF Tree e J48) e três MDE (Aster GDEM, SRTM e SRTM V3) para o MDS a nível semidetalhado, em situações em que o principal fator diferenciador entre os tipos de solo é o relevo. O uso do MDE Aster GDEM e árvore de decisão com algoritmo J48, Simple Tree e BF Tree foram os que produziram modelos de árvore de decisão capazes de produzir mapas de solo com maior similaridade ao mapa de referência.

Palavras-chave: solos, pedologia, mapa de solos, mineração de dados.

ABSTRACT

Digital soil mapping (DSM) has been shown to be feasible to use in soil survey. Although several methods have been exploited, there is a lack in defining methodologies for doing DSM. This study tests five decision trees algorithms that have been identified as suitable (Simple Chart, Random Tree, REP Tree, BF Tree, and J48) and three digital elevation models (AsterGDEM, SRTM and SRTM V3) for DSM at semidetalled level in situations where the main differentiating factor between soil types is the relief. The use of MDE Aster GDEM and decision tree algorithms J48, Simple Tree e BF Tree produced decision tree models capable of produce soil maps with larger accuracy related to reference soil maps.

Key words: soils, pedology, soil survey, data mining.

INTRODUÇÃO

Diversos estudos têm avaliado a potencialidade do mapeamento digital de solos (MDS) (ZHU, 1997; LAGACHERIE & HOLMES, 1997; BUI & MORAN, 2001; GIASSON et al., 2006; FIGUEIREDO et al., 2008; TEN CATEN et al., 2011; SARMENTO et al., 2012). Esses estudos têm tido caráter exploratório e têm identificado e avaliado métodos de predição da ocorrência de classes de solos em áreas não mapeadas, a partir da calibração dos modelos de predição através da reprodução de mapeamento de solos, feito por métodos tradicionais de levantamento.

O MDS é baseado na geração de sistemas de informações que permitem estabelecer relações matemáticas entre variáveis ambientais e solos e, dessa forma, prever a distribuição espacial das classes ou propriedades de solos. Recentemente, o uso de árvores de decisão tem se destacado por apresentar bons resultados no MDS (LAGACHERIE & HOLMES, 1997; FIGUEIREDO et al., 2008; GIASSON et al., 2011). Diversos estudos nacionais têm testado o uso de árvores de decisão no MDS, como o de COELHO & GIASSON (2010), que concluíram que árvores de decisão apresentam maior capacidade de discriminação de classes de solos do que regressões logísticas, destacando que os diversos algoritmos de árvore de decisão testados produziram mapas de solos com diferentes acurácias. Da mesma forma,

^IPrograma de Pós-graduação em Ciência do Solo, Departamento de Solos, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 7712, 91540-000, Porto Alegre, RS, Brasil. E-mail: giasson@ufrgs.br. *Autor para correspondência.

^{II}Department of Soil Science, University of Wisconsin-Madison, Madison, Estados Unidos.

GIASSON et al. (2011) testaram cinco algoritmos de árvores de decisão para o mapeamento digital de solos e indicaram que existem diferenças de desempenho entre os algoritmos de árvore de decisão que devem ser mais intensamente avaliadas no planejamento do MDS. A existência desses estudos indicando que o uso de algoritmos de árvore de decisão é adequado para o MDS ainda não é suficiente para justificar a seleção de um algoritmo com vistas à extrapolação e expansão da cobertura de mapas de solos para outras áreas circunjacentes não mapeadas.

Dado que existe crescente disponibilidade de diferentes fontes de informação sobre a elevação, outro aspecto metodológico que merece ser avaliado é o uso dos diferentes Modelos Digitais de Elevação (MDE) no MDS, já que a maioria das variáveis preditoras de ocorrência de classes de solos são derivadas de MDE. O estudo de CHAGAS et al. (2010) avaliou os MDE Aster e SRTM para aplicação em MDS, porém sem a avaliação da capacidade de predição dos modelos gerados para o MDS, pois foram estudadas somente características intrínsecas desses MDE. Como a relação entre o MDE, as variáveis do terreno e a ocorrência de solos podem variar em áreas com diferentes fisiografias, é necessário seu teste efetivamente avaliando a qualidade dos modelos através da quantificação da sua capacidade de predição de classes de solos em situações ambientalmente diversas.

O objetivo deste estudo foi testar o uso de três diferentes MDE e de cinco algoritmos árvores de decisão para a reprodução de mapa semidetalhado de solos, a partir de variáveis geomorfométricas de bacia hidrográfica previamente mapeada ao nível de semidetalhe por métodos tradicionais, localizada em área basáltica no noroeste do Estado do Rio Grande do Sul.

MATERIAL E MÉTODOS

A área de estudo é a bacia do Lajeado Grande, situada na porção noroeste do Estado do Rio Grande do Sul, Brasil. Essa área foi escolhida por bem representar uma extensa região fisiográfica do Estado do Rio Grande do Sul e por dispor de levantamento de solos na escala proposta para o estudo (KÄMPF et al., 2004) e adequada para planejamento regional e municipal. Ocupa uma área equivalente a 532km² e apresenta três fases principais de relevo: plano a suavemente ondulado, ondulado e fortemente ondulado a montanhoso. Sendo as rochas basálticas o material de origem comum a todos os solos, o

principal fator diferenciador entre os tipos de solo é o relevo, condicionando os fluxos de água, os processos erosivos e o acúmulo de materiais. O mapeamento semidetalhado de solos original foi executado na escala 1:50.000 e apresenta distribuição de classes de solos conforme a tabela 1.

Selecionados devido a sua disponibilidade global e sua gratuidade, os MDE utilizados para geração das variáveis geomorfométricas foram o ASTER GDEM (ABRAMS et al., 1999), com resolução de 30m, o SRTM (*Shuttle Radar Topography Mission*) (RABUS et al., 2003), com resolução de 3 arcos-segundos (90m) e o SRTM V3 (JARVIS et al., 2006), com resolução de 1 arco-segundo (30m). A partir de cada um dos MDE, no programa ArcGis 9.2 (ESRI, 2006), foram geradas sete variáveis geomorfométricas: a) elevação, b) declividade, c) curvatura, d) comprimento do fluxo a jusante (*downstream*), e) comprimento do fluxo a montante (*upstream*), f) acúmulo do fluxo e g) índice de umidade topográfica (BEVEN & KIRKBY, 1979), além da distância dos rios, calculada com uso da rede hidrográfica da base cartográfica vetorial 1:50.000 (HASENACK & WEBER, 2010).

As amostras para treinamento dos algoritmos foram geradas de maneira aleatória e corresponderam a um ponto amostral a cada 10ha (correspondente à Área Mínima Mapeável para a escala 1:50.000), totalizando 5320 pontos. Em cada ponto, foram registrados valores para todas as variáveis preditoras e a classe de solo ocorrente. Para a geração das árvores de decisão, as informações amostradas foram tabeladas e analisadas estatisticamente no aplicativo computacional Weka 3.5.8 (HALL et al., 2009). Para selecionar as variáveis que possuem maior poder de discriminação entre as classes de solo, foi utilizado método de seleção de atributos baseado em correlação (*Correlation-based Feature Subset Selection*) (HALL, 1999), que avalia quais atributos deve-se utilizar na predição através da avaliação da habilidade individual de predição de cada atributo e do grau de redundância entre eles.

Foi testada a predição de ocorrência de classe de solo utilizando árvores de decisão, sendo os algoritmos de árvores de decisão testados o J48, o Simple Chart, o Random Tree, o REP Tree e o BF Tree. O algoritmo J48 é uma implementação em Java do algoritmo C4.5 (QUINLAN, 1993) no programa Weka (HALL et al., 2009), que gera árvores de classificação a partir de um conjunto de dados de treinamento, sendo que, a cada nó, o algoritmo escolhe um atributo que mais eficientemente subdivide o

Tabela 1 - Legenda de identificação das unidades de mapeamento de solos mapeados na Bacia do Lajeado Grande (KÄMPF et al., 2004).

Designação da UM	Composição da UM (EMBRAPA, 2006)	Área (km ²)	Área (%)
LV1	Latossolo Vermelho distroférico típico.	144,9	27,2
LV2	Associação Latossolo Vermelho distroférico típico e Neossolo Regolítico eutrófico típico.	31,2	5,9
NV1	Nitossolo Vermelho distroférico típico.	2,4	0,4
NV2	Associação Nitossolo Vermelho distroférico típico e Neossolo Regolítico eutrófico típico.	9,0	1,7
M1	Chernossolo Argilúvico férrico típico.	3,6	0,7
M2	Associação Chernossolo Argilúvico férrico típico e Neossolo Regolítico eutrófico típico.	3,9	0,7
RR1	Neossolo Regolítico eutrófico típico.	161,8	30,4
RR2	Associação Neossolo Regolítico eutrófico típico e Neossolo Litólico Neossolo Litólico eutrófico típico, relevo forte ondulado.	125,9	23,7
RR3	Associação Neossolo Regolítico eutrófico típico e Latossolo Vermelho distroférico típico.	28,1	5,3
RR4	Associação Neossolo Regolítico eutrófico típico e Chernossolo Argilúvico férrico típico.	4,7	0,9
RR5	Associação Neossolo Regolítico eutrófico típico, Cambissolo Háplico Ta eutrófico típico e Nitossolo Vermelho distroférico típico.	2,7	0,5
RR6	Associação Neossolo Regolítico eutrófico típico e aforamentos rochosos.	2,6	0,5
RL	Associação Neossolo Litólico eutrófico típico e Neossolo Regolítico eutrófico típico.	2,2	0,4
G	Gleissolos Melânico eutrófico típico.	8,7	1,6

conjunto das amostras em subconjuntos homogêneos e caracterizados por sua classe. O critério é o ganho de informação obtida na escolha do atributo para subdivisão (QUINLAN, 1993; HALL et al., 2009). O algoritmo Random Tree usa uma classificação para a construção de árvore que considera K atributos escolhidos aleatoriamente em cada nó e não realiza a poda das árvores de decisão. O algoritmo REP Tree, por sua vez, constrói uma árvore de decisão usando informação de ganho e variância e poda a árvore usando poda de erro reduzido. O algoritmo Simple Cart usa classificação para implementação de poda com mínimo custo em complexidade e, quando trabalha com valores faltantes, usa o método de instâncias fracionadas. O algoritmo BF Tree usa classificação para construção do primeiro

melhor classificador de árvore de decisão através de divisão binária para atributos numéricos e nominais (STRATEBI OPEN BUSINESS INTELLIGENCE, 2010). A geração dos modelos de árvore de decisão utilizou o método de tabulação cruzada, disponível no programa Weka (*10 fold cross-tabulation*).

A acurácia dos modelos preditores foi avaliada no programa Weka, pela comparação das estimativas dos modelos na forma de predição de ocorrência de UM de solos com as UM de solos mapeadas no mapa original de solos e coletados nos 5320 pontos amostrais. Para essa avaliação, utilizaram-se matrizes de erro (CONGALTON, 1991), que expressam a combinação entre as UM de solos preditas por MDS e as UM no mapa de solos de referência, possibilitando a análise de indicadores de

acurácia. Os indicadores de acurácia considerados na comparação dos métodos de predição de ocorrência de classes de solos foram: a) acurácia global (AG), obtida pela divisão do número de pontos corretamente classificados com o total de pontos; e b) índice Kappa (COHEN, 1960), que mede a concordância entre as UM de solos preditas e as no mapa de referência, excluindo o efeito da casualidade.

RESULTADOS E DISCUSSÃO

As variáveis selecionadas para integrar os modelos de predição no programa Weka variaram com o tipo de MDE testado: as variáveis declividade e elevação foram selecionadas em todas as situações, sendo adicionalmente selecionados o comprimento do fluxo (quando testado o Aster GDEM) ou a curvatura (quando testados SRTM V3 e SRTM), o que demonstra que o relevo é representado diferentemente por cada um desses MDE. Os indicadores do desempenho dos

algoritmos de árvores de decisão são apresentados na tabela 2, juntamente com a representação do tamanho das árvores de decisão geradas (número de folhas finais) e do número de classes de solos preditas.

Os resultados de acurácia apresentados referem-se à concordância entre os mapas gerados e o mapa original de solos, produzido por métodos tradicionais de levantamento de solos, assumindo-se que ele seja um referencial de boa qualidade, o que nem sempre é verdadeiro e que poderia ser conhecido se viesse acompanhado por indicadores de acurácia por ocasião de sua publicação. Como não existiu avaliação da verdade de campo para nenhum dos mapas, tanto o mapa original de referência como para os gerados por MDS, não foi possível avaliar qual dos mapas melhor representa a distribuição espacial dos solos. Assim, destaca-se a necessidade de que, em estudos futuros, realize-se a avaliação dos mapas produzidos pela verdade de campo. No presente estudo, a verdade de campo não foi avaliada para

Tabela 2 - Resultados da análise das árvores de decisão em função do tipo de MDE e do algoritmo de árvore de decisão.

MDE	Algoritmo	AG	Kappa	t	UM
Aster GDEM	Simple chart	57,3	0,42	17	4
	Random tree	55,3	0,39	93	5
	REP tree	55,2	0,39	19	4
	BF tree	57,2	0,42	31	5
	J48	56,9	0,41	23	5
	Média	56,38	0,406	36,6	4,8
SRTM 30m	Simple chart	54	0,37	5	4
	Random tree	50	0,31	93	4
	REP tree	54	0,36	17	3
	BF tree	53,8	0,36	15	4
	J48	54,5	0,38	23	5
	Média	53,26	0,356	30,6	4
SRTM 90m	Simple chart	54,6	0,38	11	4
	Random tree	52,6	0,35	11	4
	REP tree	53,5	0,36	19	4
	BF tree	54,6	0,38	25	4
	J48	54,6	0,38	29	4
	Média	53,98	0,37	19	4
Média	Simple chart	55,3	0,39	11,0	4,3
	Random tree	52,6	0,35	65,7	4,3
	REP tree	54,2	0,37	18,3	3,7
	BF tree	55,2	0,39	23,7	4,3
	J48	55,3	0,39	25,0	4,7
	Média	54,5	0,38	28,7	4,3

AG = acurácia geral; Kappa = índice Kappa; t = tamanho da árvore de decisão expresso em número de folhas finais; UM = número de unidades de mapeamento de solos que o algoritmo foi capaz de prever.

os mapas gerados por causa de limitações logísticas e financeiras de se ir a campo descrever perfis para validar os mapas gerados.

Nenhuma combinação de algoritmo de árvore de decisão e MDE foi capaz de modelar a ocorrência de todas as UM de solos presentes no mapa original de referência (15 UM), dada a existência de UM que ocorre em áreas muito pequenas na área de estudo. Das 15 UM presentes no mapa de solos de referência, somente de 3 a 5 foram preditas, sendo estas as UM que ocorrem em áreas mais extensas (LV1, RR1, RR2, LV2 e RR3, em ordem decrescente de extensão).

A utilização do MDE Aster GDEM gerou predições de ocorrência de classes de solos com acurácia levemente superior aos demais, com mais altos valores de AG e índice kappa para todos os algoritmos utilizados. Esse resultado verifica-se provavelmente porque o MDE ASTER (resolução espacial de 30 m) apresenta informações mais detalhadas, referentes à variação do relevo que o MDE SRTM (resolução espacial de 90m) e mesmo do SRTM v3, que é uma modificação do SRTM. Esse resultado é coerente com resultados de diversos estudos, como de CAMARGO et al. (2009), que concluíram que o uso do ASTER é viável para a semi-automação de procedimentos relacionados ao mapeamento de unidades de relevo. Todavia, esse resultado contradiz a recomendação de Guth (2010) de usar preferencialmente o SRTM, dado que em seu estudo não foram encontradas diferenças significativas entre SRTM, SRTM V3 e Aster GDEM e que este pode apresentar anomalias em 20% das vezes. Dado que a acurácia dos modelos gerados com Aster GDEM são apenas levemente superiores às dos demais MDE, não tem-se suficientes evidências para recomendar o uso de nenhum dos MDE.

Todos os algoritmos utilizados apresentaram capacidade semelhante de predizer a ocorrência das UM de solos, variando de 50% a 57,3% de acurácia geral, valores esses que estão na mesma faixa de grandeza da acurácia obtida em estudos recentes na mesma área de estudos (GIASSON et al., 2011). Esses valores não muito elevados da acurácia podem significar que as variáveis utilizadas na modelagem não são suficientes para explicar completamente a distribuição dos solos na paisagem, dado que essas foram derivadas de apenas um fator de formação do solo (relevo). Na média dos diversos MDE, os modelos usando algoritmo Simple Chart, J48 e BF Tree produziram árvores de decisão com capacidade ligeiramente superior de gerar mapas semelhantes aos mapas de referência, apresentando mais altos valores de AG média (55,3% para Simple Chart e J48 e

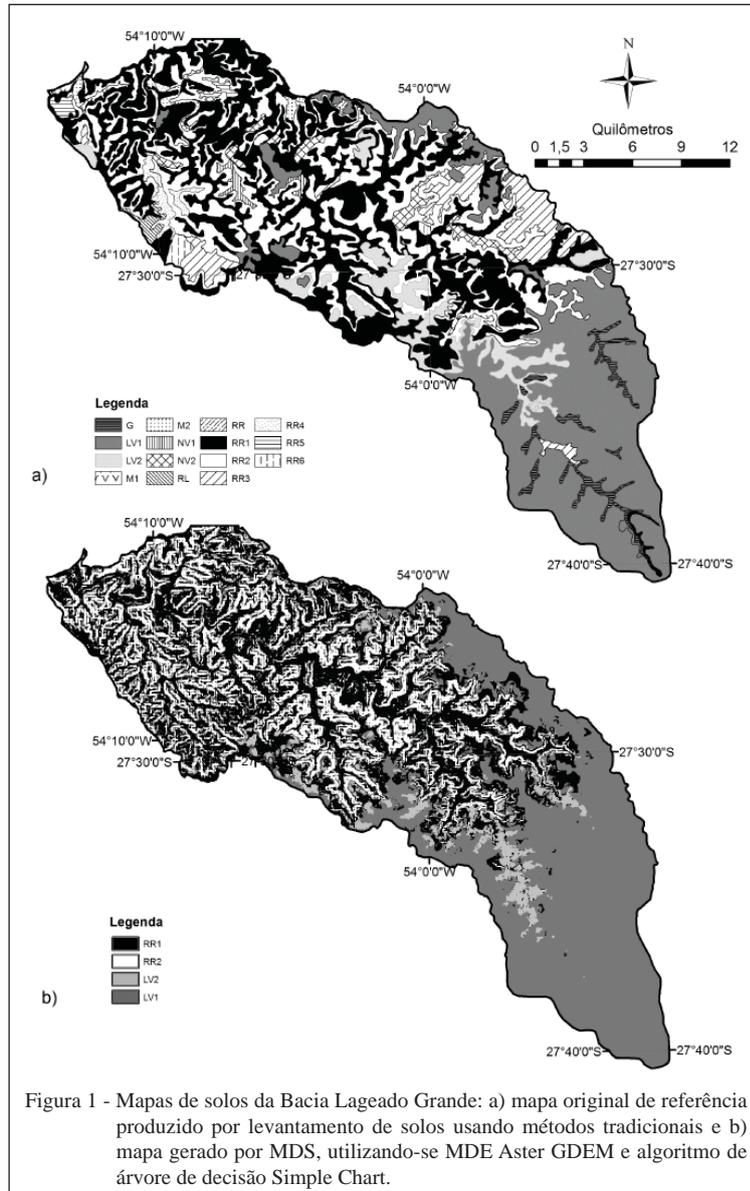
55,2% para BF Tree) e índice kappa médio (0,39 para os três algoritmos). No geral, as árvores de decisão geradas com Simple Chart apresentam árvores de decisão mais simples (menor tamanho), o que é uma vantagem, dado que, atualmente, a implementação das regras dos modelos de árvore de decisão gerados no programa Weka tem que ser realizada manualmente para a confecção do mapa de solos em sistemas de informações geográficas e que a geração de árvores menores facilita a sua implementação e reduz a chance de introdução de erros. A partir desses resultados e a fim de ilustrar graficamente os mapas gerados pelo MDS, foi gerado mapa, utilizando MDE Aster GDEM e algoritmo de árvore de decisão Simple Chart, apresentado na figura 1. A observação desta figura permite perceber a espacialização das UM onde houve mais concordância entre os dois mapas, como a porção sudeste da bacia, onde predomina a UM LV1 (Latosolo Vermelho distroférrico típico), em relevo suave ondulado e ao redor dos eixos centrais de drenagem dos principais rios da bacia, onde, em relevo forte ondulado a montanhoso, ocorre centralmente a UM RR1 (Neossolo Regolítico eutrófico típico), que vem cercada por áreas da UM RR2 (Associação Neossolo Regolítico eutrófico típico e Neossolo Litólico eutrófico típico, relevo forte ondulado).

CONCLUSÃO

O mapeamento digital de solos utilizando MDE Aster GDEM e árvore de decisão com algoritmo J48, Simple Tree e BF Tree foi o que produziu modelos de árvore de decisão capazes de produzir mapas de solos com maior semelhança com o mapa de referência. A comparação dos vários MDE e dos vários algoritmos de árvore de decisão não foi conclusiva, pois as diferenças entre os mesmos foram pequenas. A avaliação dos resultados e metodologia deste estudo indica que, em estudos futuros, deve-se utilizar mapas de solos tradicionais de referência, com qualidade conhecida. Além disso, também indica que os mapas gerados por MDS devem ser avaliados através da verdade de campo, o que poderá dar indicações mais úteis sobre a qualidade dos.

AGRADECIMENTOS

Este estudo foi financiado por Bolsa de Produtividade em Pesquisa concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) ao primeiro autor e bolsas de doutorado da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e CNPq, respectivamente, para o quarto autor e o quinto autor.



REFERÊNCIAS

ABRAMS, M. et al. **ASTER users' handbook**. Pasadena: JPL, 1999. 93p.

BEVEN, K.; KIRKBY, N. A physically based variable contributing area model of basin hydrology. *Hydrological sciences. Bulletin des Sciences Hydrologiques*, v.24, p.43-69, 1979. Disponível em: <<http://dx.doi.org/10.1080/02626667909491834>>. Acesso em: 03 ago. 2011. doi: 10.1080/02626667909491834.

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

CAMARGO, F.F. et al. Abordagens cognitivas para a identificação de unidades geomorfológicas por meio de dados ASTER/Terra. *Revista Brasileira de Geociências*, v.39, n.2, p.276-288, 2009.

Disponível em: <http://ppegeo.igc.usp.br/scielo.php?script=sci_arttext&pid=S0375-75362009000200008&lng=en&nrm=iso>. Acesso em: 05 nov. 2012.

CHAGAS, C. da S. et al. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. *Pesquisa Agropecuária Brasileira*, v.45, n.5, p.497-507, 2010. Disponível em: <<http://dx.doi.org/10.1590/S1415-43662009000100007>>. Acesso em: 25 maio, 2011. doi: 10.1590/S1415-43662009000100007.

COELHO, F.F.; GIASSON, E. Comparison of methods for digital soil mapping using a geographical information system. *Ciência Rural*, v.40, n.10, 2010. Disponível em: <<http://dx.doi.org/10.1590/S0103-84782010005000156>>. Acesso em: 05 nov. 2012. doi: 10.1590/S0103-84782010005000156.

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

- Disponível em: <<http://dx.doi.org/10.1177/001316446002000104>>. Acesso em: 03 ago. 2011. doi: 10.1177/001316446002000104.
- CONGALTON, R.G. A review of assessing the accuracy of classification of remotely sensed data. **Remote Sensing of Environment**, v.37, p.35-46, 1991. Disponível em: <[http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B)>. Acesso em: 03 ago. 2011. doi: 10.1016/0034-4257(91)90048-B.
- EMBRAPA. **Sistema brasileiro de classificação de solos**. 2.ed. Rio de Janeiro: Centro Nacional de Pesquisa de Solos, 2006. 306p.
- ESRI. **ArcGis 9.2**. Redland, California, 2006. Software.
- FIGUEIREDO, S.R. et al. Uso de regressões logísticas múltiplas para mapeamento digital de solos no Planalto Médio do RS. **Revista Brasileira de Ciência do Solo**, v.32, p.2779-2785, 2008. Disponível em: <<http://dx.doi.org/10.1590/S0100-06832008000700023>>. Acesso em: 05 nov. 2012. doi: 10.1590/S0100-06832008000700023.
- GIASSON, E. et al. Digital soil mapping using multiple logistic regressions on terrain parameters in Southern Brazil. **Scientia Agricola**, v.63, n.3, p.262-268, 2006. Disponível em: <<http://dx.doi.org/10.1590/S0103-90162006000300008>>. Acesso em: 03 ago. 2011. doi: 10.1590/S0103-90162006000300008.
- GIASSON, E. et al. Decision trees for digital soil mapping on subtropical basaltic steepplands. **Scientia Agricola**, v.68, n.2, p.167-174, 2011. Disponível em: <<http://dx.doi.org/10.1590/S0103-90162011000200006>>. Acesso em: 03 ago. 2011. doi: 10.1590/S0103-90162011000200006.
- GUTH, P.L. **Geomorphometric comparison of ASTER GDEM and SRTM**. A special joint symposium of ISPRS Technical Commission IV & AutoCarto in conjunction with ASPRS/CaGIS 2010. Disponível em: <<http://www.isprs.org/proceedings/XXXVIII/part4/files/Guth.pdf>>. Acesso em 03 maio 2013.
- LAGACHERIE, P.; HOLMES, S. Addressing geographical data errors in a classification tree soil unit prediction. **International Journal of Geographical Information Science**, v.11, p.183-198, 1997.
- HALL, M.A. **Correlation-based feature subset selection for machine learning**. 1999. 178f. Tese - Department of Computer Science, University of Waikato.
- HALL, M. et al. **The WEKA Data Mining Software: an update**; SIGKDD Explorations Newsletter, v.11, n.1, p.10-18, 2009.
- HASENACK, H.; WEBER, E. (Org.). **Base cartográfica vetorial contínua do Rio Grande do Sul - escala 1:50.000**. Porto Alegre: UFRGS-IB-Centro de Ecologia, 2010. 1 DVD-ROM. (Série Geoprocessamento, 3).
- JARVIS, A. et al. **Hole-filled seamless SRTM data V3**. International Centre for Tropical Agriculture (CIAT), 2006. Acessado em 10 ago. 2012. Disponível em: <<http://srtm.csi.cgiar.org/>>. Acesso em: 03 ago. 2011.
- KÄMPF, N. et al. **Levantamento pedológico e análise qualitativa do potencial de uso dos solos para o descarte de dejetos suínos da microbacia do Lajeado Grande**. Porto Alegre: Secretaria de Estado do Meio Ambiente do RS, Programa Nacional do Meio Ambiente II, 2004. 122p.
- LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P. et al. (Ed). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007. p.3-24.
- QUINLAN, J.R. **C4.5: programs for machine learning**. Sydney, Austrália: Morgan Kaufmann Publishers, 1993. 302 p.
- RABUS, B. et al. The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar. **Journal of Photogrammetry & Remote Sensing**, v.57, p.241-262, 2003. Disponível em: <[http://dx.doi.org/10.1016/S0924-2716\(02\)00124-7](http://dx.doi.org/10.1016/S0924-2716(02)00124-7)>. Acesso em: 03 ago. 2011. doi:10.1016/S0924-2716(02)00124-7.
- SARMENTO, E.C. et al. Predição de ordens de solos com alta resolução espacial: resposta de diferentes classificadores à densidade de amostragem. **Pesquisa Agropecuária Brasileira**, v.47, n.9, p.1395-1403, 2012. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-204X2012000900025&script=sci_arttext>. Acesso em: 19 nov. 2012.
- STRATEBI OPEN BUSINESS INTELLIGENCE. **Comparativa de algoritmos de ferramentas de data mining**. 2010. Disponível em: <http://www.stratebi.es/todobi/ago10/Algoritmos-Herramientas_Data_Mining.pdf>. Acesso em: 19 jun. 2012.
- TEN CATEN, A. et al. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. **Revista Brasileira de Ciência do Solo**, v.35, n.1, p.53-62. Disponível em: <<http://www.scielo.br/pdf/rbcs/v35n1/a05v35n1.pdf>>. Acesso em 03 ago. 2012.
- ZHU, A.X. 1997. A similarity model for representing soil spatial information. **Geoderma**, v. 77, p. 217-242. Disponível em: <[http://dx.doi.org/10.1016/S0016-7061\(97\)00023-2](http://dx.doi.org/10.1016/S0016-7061(97)00023-2)>. Acesso em 03 ago. 2012.