



Revista Brasileira de Ciência do Solo

ISSN: 0100-0683

revista@sbc.org.br

Sociedade Brasileira de Ciência do Solo
Brasil

Caten, Alexandre ten; Diniz Dalmolin, Ricardo Simão; Araújo Pedron, Fabrício; Mendonça-Santos, Maria de Lourdes

Regressões Logísticas Múltiplas: fatores que influenciam sua aplicação na predição de classes de solos

Revista Brasileira de Ciência do Solo, vol. 35, núm. 1, enero-febrero, 2011, pp. 53-62

Sociedade Brasileira de Ciência do Solo

Viçosa, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=180219142005>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

1.3 - Pedometria

REGRESSÕES LOGÍSTICAS MÚLTIPLAS: FATORES QUE INFLUENCIAM SUA APLICAÇÃO NA PREDIÇÃO DE CLASSES DE SOLOS⁽¹⁾

Alexandre ten Caten⁽²⁾, Ricardo Simão Diniz Dalmolin⁽³⁾, Fabrício Araújo Pedron⁽⁴⁾ & Maria de Lourdes Mendonça-Santos⁽⁵⁾

RESUMO

Métodos mais eficazes para determinação do padrão de distribuição de classes de solo na paisagem precisam ser avaliados visando suprir a demanda por mapas de solo em escalas regional e global. Neste estudo, Regressões Logísticas Múltiplas foram utilizadas como modelos preditores em uma aplicação de Mapeamento Digital de Solos. Os modelos foram gerados utilizando um mapa de solos existente como variável dependente e atributos de terreno como variáveis independentes, o que possibilitou determinar a probabilidade de encontrar classes de solo na paisagem no primeiro e no segundo nível categórico do SiBCS. A qualidade dos mapas preditos foi verificada por meio da matriz de contingência. A classe dos Argissolos foi predita corretamente, em relação ao mapa original, em aproximadamente 85 %. As classes de solos hidromórficos (Planossolos e Gleissolos) foram preditas corretamente em 75 %. Houve confundimento dos modelos para as classes que ocupam posições muito semelhantes na paisagem. Foi verificado também que classes de solo pouco representativas na paisagem não são adequadamente espacializadas em razão da sensibilidade dos modelos logísticos à proporção relativa das amostras usadas para treinar os modelos.

Termos de indexação: mapeamento digital de solos, pedometria, modelos lineares generalizados.

-
- ⁽¹⁾ Parte da Dissertação de Mestrado do primeiro autor apresentada ao Programa de Pós-Graduação em Ciência do Solo da Universidade Federal de Santa Maria. Estudo financiado pela CAPES e pelo CNPq. Recebido para publicação em janeiro de 2010 e aprovado em dezembro de 2010.
- ⁽²⁾ Doutorando do Programa de Pós-Graduação em Ciência do Solo da UFSM, Professor do Instituto Federal Farroupilha Campus Júlio de Castilhos. São João do Barro Preto, s/n, CEP:98130-000 Júlio de Castilhos (RS). E-mail: acaten@yahoo.com.br
- ⁽³⁾ Professor Associado do Departamento de Solos, Centro de Ciências Rurais, Universidade Federal de Santa Maria – UFSM. Av. Roraima 1000, Camobi, CEP 97105-900 Santa Maria (RS). Bolsista CNPq. E-mail: dalmolinr@pq.cnpq.br
- ⁽⁴⁾ Professor Adjunto do Departamento de Solos, Centro de Ciências Rurais, UFSM. E-mail: fapedron@gmail.com
- ⁽⁵⁾ Pesquisadora da Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Solos – EMBRAPA/CNPQ. Rua Jardim Botânico, 1024, CEP 22460-000 Rio de Janeiro (RJ). E-mail: loumendonca@cnpq.embrapa.br

SUMMARY: MULTIPLE LOGISTIC REGRESSIONS: CONTROLLING FACTORS IN APPLICATIONS TO SOIL CLASS PREDICTION

More effective methodologies to determine the soil class distribution must be evaluated in order to meet the demand for soil maps at regional and global scales. In this study, logistic regressions were used as predictive models in an application of Digital Soil Mapping. The models were derived from an existing soil map as dependent variable and terrain attributes as independent variables. The probability of finding soil classes in the landscape at the 1st and 2nd Categorical Level of the Brazilian System of Soil Classification (SiBCS) was determined. The quality of the predicted map was tested using a contingency matrix. Approximately 85 % of the Acrisols (Argissolos) were correctly predicted, in relation to the original map. Of the hydromorphic soils, 75 % were correctly predicted. The prediction was inaccurate for classes in very similar positions in the landscape. It was also found that the non-representative soil classes of the landscape were not properly spatialized, due to sensitivity of the logistic models to the relative proportion of the samples used to adjust the models.

Index terms: digital soil mapping, pedometry, generalized linear models.

INTRODUÇÃO

As demandas da sociedade por informações sobre os solos têm aumentado muito devido à renovação do interesse global pela produção agrícola e pelas questões de sustentabilidade ambiental e de mudanças climáticas (Hartemink & McBratney, 2008), entre outras demandas emergentes, como produção de bioenergia e geotecnia. Esses enfoques foram adicionados aos convencionais interesses em fertilidade e manejo do solo para uso agrícola e zoneamentos agroecológicos, atraindo assim novos usuários na busca de informações sobre o solo (Mermut & Eswaran, 2001).

Diante dessas novas demandas, os dados e informações de solos disponíveis precisam ser organizados em bancos de dados, a fim de facilitar sua manipulação e uso, servindo inclusive para a tomada de decisão em relação à necessidade de novas amostragens. Nesse sentido, além de bancos de dados, outras técnicas disponíveis e métodos quantitativos têm sido utilizados para predição espacial e mapeamento digital de solos (classes e propriedades), constituindo o que hoje é aceito pela *International Union of Soil Sciences* (IUSS) como Pedometria (McBratney et al., 2003).

A predição e o Mapeamento Digital de Solos (MDS) – *Digital Soil Mapping* – foram definidos por Lagacherie & McBratney (2007) como “a criação e a população de sistemas de informação espacial de solos por meio de modelos numéricos para inferir as variações espaciais e temporais de classes de solo e suas propriedades, a partir de observações, conhecimento e de dados de variáveis ambientais relacionados”. A principal aplicação dessa abordagem é a predição por meio de equações matemáticas, de classes e propriedades de solos e o mapeamento digital

dos resultados de forma contínua e espacial, criando a possibilidade de organizar um amplo conjunto de dados para análise e interpretações em qualquer época, não sendo o mapa o único produto (McBratney et al., 2003).

De acordo com Bui et al. (2006), técnicas de mineração de dados têm sido extensivamente utilizadas na busca de identificar e extrair conhecimento de mapas de solos existentes. Por meio de uma abordagem de engenharia reversa (Voltz et al., 1997), o conhecimento pedológico contido nos mapas de solos é extraído e formalizado, sendo essas informações usadas para gerar os modelos preditivos que serão aplicados às áreas onde não existem levantamentos de solos, mas que apresentam relação solo-paisagem semelhante.

Em casos nos quais o resultado de uma inferência pode ser dado sob muitas categorias ou politômica (classes de solo), uma alternativa é trabalhar com a probabilidade de ocorrência de cada uma das categorias, utilizando-se a Regressão Logística Múltipla (RLM). De acordo com Chatterjee & Hadi (2006), a RLM não tem qualquer exigência para sua aplicação no que diz respeito à distribuição das variáveis explicativas, não havendo necessidade de ter distribuição normal, correlação linear, medidas em mesma escala ou homogeneidade de variância. As variáveis explicativas podem ser, ainda, uma mistura de dados contínuos, discretos ou binários.

A aplicação de RLM envolve a função logarítmica ou *logit*, a qual é definida como a razão entre a probabilidade de o evento ocorrer e a probabilidade de esse evento não ocorrer (*odds* ou razão de chance); ou, ainda, exemplificando com classes de solo, a *logit* é a função logarítmica da razão entre a probabilidade (p_i) de um pixel ser membro de uma classe de solo “i” e a probabilidade de que ele não seja ($1 - p_i$). Seu valor pode ser determinado por meio da regressão:

$$\text{logit}_i = \ln\left(\frac{p_i}{(1-p_i)}\right) = a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (1)$$

A equação demonstra como calcular o *logit* de uma categoria predita a partir de um conjunto de variáveis ($X_{i...k}$). O *logit* é, por conseguinte, o logaritmo natural de uma razão entre probabilidades. O valor de “a” indica o intercepto da regressão e os valores de “b” são os coeficientes de cada variável explicativa (Chatterjee & Hadi, 2006).

A aplicação de RLM ao mapeamento do solo tem recebido pouco enfoque da pesquisa, conforme constatado por McBratney et al. (2003) e Kempen et al. (2009). Apesar disso, trabalhos recentes utilizando RLM para o mapeamento do solo podem ser encontrados na literatura internacional (Hengl et al., 2007; Debella-Gilo & Etzelmüller, 2009). No Brasil, constatou-se o uso de RLM em estudos de mapeamentos de solos semidetalhados realizados por Giasson et al. (2006) e Figueiredo et al. (2008). Percebe-se, portanto, que o número de trabalhos é pequeno e existe ainda uma grande lacuna no conhecimento para a aplicação dessas técnicas de modelagem linear generalizada.

O objetivo deste trabalho foi avaliar a utilização de RLM como modelo preditor aplicado ao MDS, a influência da proporção relativa das amostras de cada classe de solo, bem como a posição na paisagem da classe a ser predita.

MATERIAL E MÉTODOS

Área de estudo

A área de estudo é o município de São Pedro do Sul, com uma superfície de 874 km², situado na região central do Estado do Rio Grande do Sul (RS) - Brasil. Essa área foi escolhida por apresentar um levantamento de solos semidetalhado na escala 1:50.000 (Klamt et al., 2001), além de apresentar grande variação de relevo por estar localizada em uma região transitória entre o Planalto Médio e Depressão Central do RS. A figura 1 mostra o mapa de solos e também a seção longitudinal do município, com a geologia e a altitude das três principais formas da paisagem que ocorrem na área de estudos.

O mapa de solos foi digitalizado utilizando-se o programa ArcGIS 9.3 (ESRI, 2008), por meio do aplicativo Editor, com auxílio do procedimento Auto-completar, o que possibilitou um delineamento sem conflitos ou lacunas entre polígonos vizinhos. Por esse método, classes de solos vizinhas têm seus limites digitalizados apenas uma única vez.

Atributos de terreno

Como forma de garantir uma mesma origem para as variáveis preditivas, foram utilizados os seguintes atributos de terreno: elevação (ELEV) – com

importante papel na definição do clima local; declividade (DECL) – a qual afeta a velocidade de fluxos superficiais e subsuperficiais; curvatura planar (CPLN) – que indica a concentração ou dispersão da água sob o relevo; curvatura de perfil (CPRF) – importante para a velocidade da água; distância à drenagem (DIST) – relacionada com a densidade da rede de drenagem; radiação relativa disponível (RADI) – representando diferentes exposições da paisagem à energia solar; área de contribuição (LNAC) – relacionada à disponibilidade de água para os processos ambientais; índice de umidade topográfica (IUT) – indicando o controle da topografia sobre a umidade do solo; e capacidade de transporte de sedimento (CTS) – caracterizando processos de erosão e deposição na paisagem (Wilson & Gallant, 2000).

Para a derivação dos atributos do terreno utilizados no estudo, visando à representação dos condicionantes da formação do solo, foi empregado o Modelo Digital de Elevação (MDE) utilizando informações do *Shuttle Radar Topographic Mission* (SRTM). Os modelos matemáticos estatísticos usados para gerar cada um dos nove atributos de terreno estão detalhadamente descritos em Wilson & Gallant (2000).

Mapa de solos digital

Os Planos de Informações (PIs) dos atributos do terreno (nove ao todo), bem como o arquivo contendo o mapa de solos de São Pedro do Sul (mapa original), foram convertidos para o formato raster constituindo um *projeto* do programa ArcGIS 9.3. Um total de 70.000 pixels (cada pixel com 50 m de resolução espacial), representando aproximadamente 20 % da área total do município, foram aleatoriamente gerados para a amostragem e tabulação dos dados de variáveis do terreno e das classes de solo. A amostragem foi executada com a função *Sample* no programa ArcGIS. A matriz de treinamento dos modelos constituiu-se de 10 colunas (nove atributos extraídos do MDE mais a classe de solo no pixel) e 70.000 linhas.

Na sequência, esses dados foram utilizados como variáveis explicativas em RLM. Como variável-resposta, utilizaram-se as classes de solo em nível de ordem e subordem – primeiro Nível Categórico (1° NC) e segundo Nível Categórico (2° NC), respectivamente – do Sistema Brasileiro de Classificação de Solos (Embrapa, 2006) a partir do mapa original (Klamt et al., 2001). Em todos os modelos *logit* testados, foram considerados apenas os coeficientes acima de 5 % de significância pelo teste de Wald (Chatterjee & Hadi, 2006).

Utilizando a função *Raster Calculator* do ArcGIS, foram processadas todas as etapas de produção dos PIs. Desse processamento originaram-se PIs de probabilidade de ocorrência para cada uma das classes de solo. O procedimento final constituiu-se da verificação do maior valor de probabilidade em cada ponto da paisagem (pixel) entre todos os PIs, originando o mapa de solos predito.

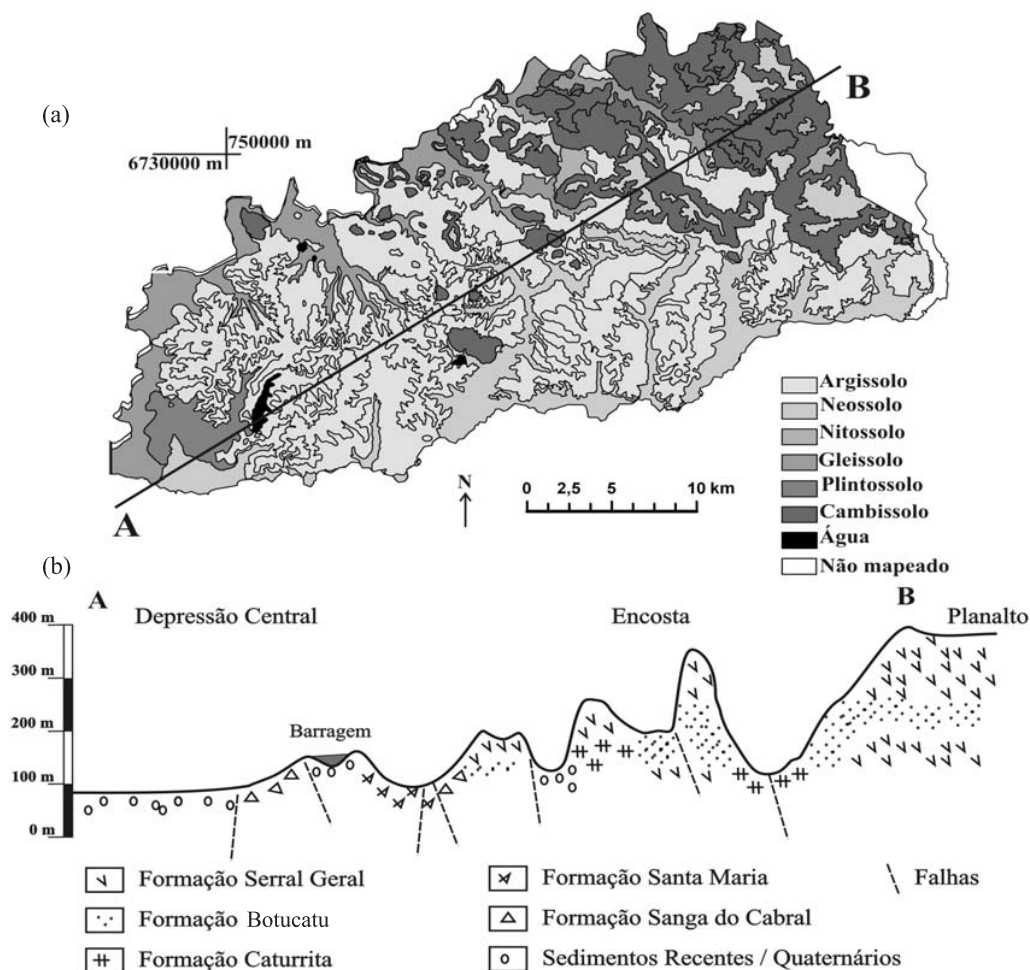


Figura 1. Mapa das ordens de solos existentes na área de estudo (Klamt et al., 2001), coordenadas E e N no sistema UTM/SAD69, fuso 21 (a); seção longitudinal do município, localizando a geologia e a altitude das três principais formas da paisagem na área de estudo (b).

A verificação da acurácia do mapa predito serviu para testar a capacidade dos modelos RLM em reproduzir o mapa original. Um novo conjunto de 7.000 pixels, correspondendo a 2 % do total da área de estudo, foi aleatoriamente gerado, servindo para a amostragem do mapa original e do mapa predito. Essas informações tabuladas, no formato ASCII, foram processadas para a geração da matriz de erros (Congalton, 1991).

Foi utilizado um conjunto mínimo de 100 amostras para cada classe a ser predita, uma vez que a estimativa dos parâmetros da função *logit* foi realizada por máxima verossimilhança. Essa regra prática foi cumprida em todos os conjuntos de dados utilizados para formulação das funções *logit*, conforme recomendação de Peng et al. (2002).

RESULTADOS E DISCUSSÃO

Modelos *logit*

Foram utilizadas como referência para as RLM as classes dos Cambissolos e dos Solos Hidromórficos (Planossolos e Gleissolos) (Quadro 1), buscando-se testar a influência dessas diferentes classes de solos no desenvolvimento dos modelos. As RLM utilizadas tiveram os interceptos (*a* da função *logit*) e coeficientes (*b* da função *logit*) conforme os dados do quadro 1. Nessa tabela, os valores não significativos não foram usados nas respectivas equações.

Os modelos *logit* gerados e apresentados no quadro 1 foram obtidos conforme exemplificados na equação 2 para a classe dos Neossolos.

$$\text{logit}_R = e^{1,087 + 0,011 \cdot \text{ELEV} - 0,123 \cdot \text{DECL} - 1,608 \cdot \text{CPLN} + 0,901 \cdot \text{CPRF} - 0,011 \cdot \text{ILUM} - 0,419 \cdot \text{LNAC} + 0,200 \cdot \text{IUT}}$$

(2)

Quadro 1. Valores dos interceptos e coeficientes no primeiro e no segundo nível categórico, gerados a partir das covariáveis preditoras (atributos de terreno) e covariáveis-resposta (classes de solo)

Preditor	1º Nível Categórico - Ordem					2º Nível Categórico - Subordem						
	Cambissolos ⁽¹⁾					Solos Hidromórficos ⁽¹⁾						
	R	SH	P	N	F	FT	PAC	PBAC	PV	PVA	RL	CX_NV
Intercepto	1,087	15,344	10,963	-1,551	20,006	4,742	-8,555	-2,838	-8,173	-7,732	-16,304	-17,080
ELEV	0,011	-0,087	-0,031	0,006	-0,111	-0,024	0,067	0,035	0,083	0,073	0,114	0,103
DECL	-0,123	-0,209	-0,221	-0,218	-0,488	-0,284	0,029	-0,101	-0,062	0,071	0,134	0,264
DIST	0,001	-0,003	-0,003	-0,001	0,005	0,003	0,002	0,003	0,003	0,003	0,003	0,003
CPLN	-1,608	3,716	-0,449	-0,287	1,598	-2,134	-3,960	-4,221	-4,887	-4,141	-5,429	-3,850
CPRF	0,901	2,756	1,869	1,697	3,495	0,763	-0,871	-0,058	-1,963	-1,539	-2,369	-3,230
RADI	-0,011	0,017	-0,002	0,008	-0,008	-0,026	-0,016	-0,017	-0,033	-0,022	-0,032	-0,020
LNAC	-0,419	-0,408	-0,092	-0,368	-0,007	0,403	0,357	0,443	0,386	0,286	-0,030	0,261
IUT	0,200	0,080	-0,153	0,187	-0,276	-0,355	-0,214	-0,346	-0,353	-0,160	0,160	0,072
CTS	-0,004	0,003	-0,010	-0,011	-0,007	-0,012	-0,048	-0,039	-0,004	-0,009	-0,006	-0,002

⁽¹⁾ Classe de referência. Valores em negrito não foram significativos a 5 % (teste de Wald). Todos os modelos apresentaram diferença significativa para o teste dos modelos apenas com o intercepto e os mesmos com as variáveis explicativas. Todas as variáveis explicativas têm relação significativa com a variável-resposta (teste χ^2); Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F), associação Cambissolo Háplico e Nitossolo Vermelho (CX_NV), Argissolo Vermelho (PV), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC), Plintossolo Argilúvico (FT), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Radiação Relativa Disponível (RADI), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

A partir dos parâmetros, para cada classe de solo, exceto a classe de referência, obteve-se um modelo *logit*, possibilitando prever a probabilidade de que determinada classe de solo possa ser encontrada em um dado local da paisagem. Sendo assim, as equações propostas, linearizaram a relação entre as covariáveis e o *logit* de cada classe de solo.

As covariáveis ELEV, DECL, RADI e IUT obtiveram todos seus parâmetros estimados significativos nos dois níveis categóricos testados. Quanto à influência de cada uma das covariáveis, Giasson et al. (2006) verificaram que os atributos do terreno mais recorrentes quando da utilização de regressões logísticas foram ELEV, DIST, IUT, curvaturas e DECL, os quais são atributos relacionados com acúmulo e dinâmica da água. Para Debella-Gilo & Etzelmüller (2009), os atributos do terreno que mais influenciaram em modelos utilizando-se de RLM foram ELEV, DIST, tempo de exposição à radiação, DECL e IUT.

A escolha das classes de referência utilizadas neste estudo não mostrou evidências de que tenha produzido diferentes resultados do ponto de vista do padrão dos mapas de probabilidade de cada classe de solo. Contudo, a influência de cada uma das classes de solo, como referência na fase de determinação dos parâmetros das equações *logit*, pode ser mais bem investigada em futuros trabalhos utilizando de RLM.

Mapas de probabilidade

A aplicação dos modelos *logit* gerou arquivos raster de probabilidade, denominados mapas de probabilidades (Figura 2). Em cada PI, os pixels representam a

probabilidade (entre 0 e 1) de se encontrar a classe de solo na paisagem, similar ao valor de pertinência das técnicas de Lógica Nebulosa (*Fuzzy Logic*) (Debella-Gilo & Etzelmüller, 2009). Os planos de informações (PIs) que representam a probabilidade de encontrar cada uma das classes de solo têm uma boa relação visual com o que se esperava encontrar pela relação solo-paisagem para a região. As maiores probabilidades de encontrar os Solos Hidromórficos (Planossolos e Gleissolos) ocorreram nos locais mais baixos (relevo plano), onde a presença de água é um importante fator no processo de pedogênese. Nos topos dos morros e na encosta do Planalto (relevo forte ondulado a montanhoso) observaram-se as maiores probabilidades de encontrar as classes dos Neossolos Litólicos e Cambissolos, onde, especialmente para o primeiro, o processo de erosão não permite o desenvolvimento de solos mais profundos. Nas áreas de colinas (relevo suave ondulado a ondulado), intermediárias entre várzeas e encostas de morros, encontraram-se as maiores probabilidades para os Argissolos que apresentam horizontes mais espessos e são bem drenados.

Esse potencial das RLM em reproduzir a relação solo-paisagem existente no mapa original utilizado para gerar os modelos havia sido reportado por Debella-Gilo & Etzelmüller (2009) em solos da Noruega. De acordo com esses autores, os mapas de probabilidade possuíam um padrão de distribuição espacial similar ao que se esperava encontrar a partir do conhecimento da relação solo-paisagem.

A reunião dos PIs de probabilidade de cada classe de solo em um único PI, contendo apenas o maior valor para cada pixel, pode ser observada no estrato da área

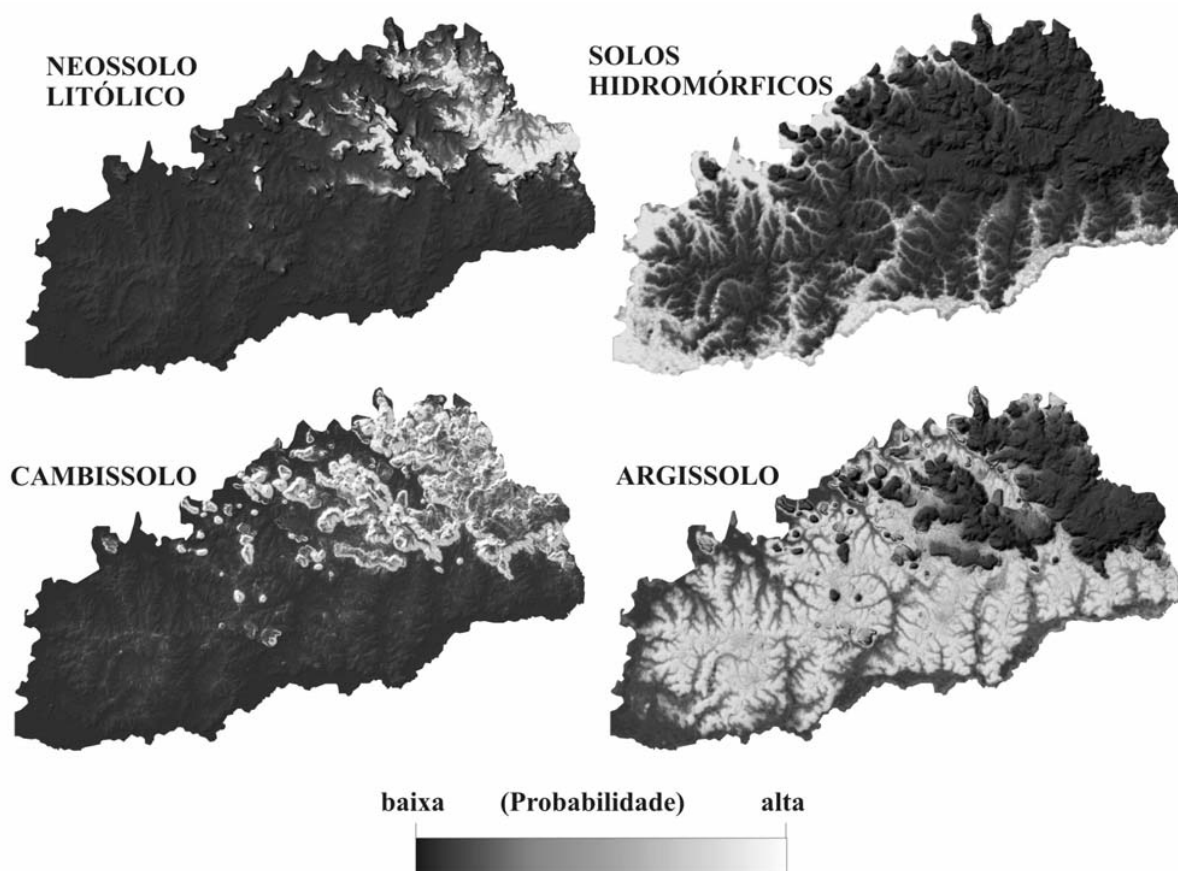


Figura 2. Mapas de probabilidade de ocorrência das classes de solos em nível de ordem na área de estudo.

de estudo da figura 3a. Os maiores valores de probabilidade estão ligados aos solos encontrados em regiões distintas, quanto aos atributos do terreno utilizados nos modelos, as quais são as regiões dos topos dos morros e encosta do planalto e as várzeas ao longo das drenagens (Figura 3a).

Os valores de probabilidades mais baixos – como, por exemplo, para a classe dos Argissolos, que têm boa representatividade na área de estudo – teriam seu valor aumentado se um número maior de covariáveis ligadas ao relevo fosse utilizado ou, ainda, se informações como a litologia ou uso da terra fossem levadas em consideração.

Como a RLM gera um mapa de probabilidade de se encontrar cada uma das classes de solo na paisagem, essa informação poderá ter uma aplicação em outros estudos que se utilizem da distribuição espacial do solo, como os que tratam de probabilidade da ocorrência de deslizamentos de terra (Ohlmacher & Davis, 2003), armazenamento de água pelo solo ou sequestro de C. Nesse caso, a informação de probabilidade de se encontrar cada uma das classes de solo na paisagem, variando entre zero e um, seria utilizada como peso ou ponderador nos modelos.

Mapas de classe de solos

A partir do PI com os maiores valores de probabilidade para todas as classes de solo, gerou-se um mapa de solos para a área do estudo (Figura 3b). Os modelos logísticos foram capazes de distinguir três posições distintas na paisagem: relevo plano, relevo suave ondulado a ondulado e relevo forte ondulado a montanhoso. As classes Cambissolos, Argissolos e os Solos Hidromórficos foram espacializadas visualmente de modo semelhante ao que se encontra no mapa original e, consequentemente, à relação solo-paisagem da área de estudo.

Figueiredo et al. (2008) estabeleceram uma comparação visual entre o mapa predito e o mapa original, tendo verificado elevada correlação entre ambos, o que lhes permitiu afirmar que os modelos logísticos têm potencial para aproximar as pedoformas preditas das pedoformas originais.

Neste estudo o mapa de solos predito não espacializou a classe dos Plintossolos, possivelmente devido à pouca representatividade dessa classe no total das amostras utilizadas nos modelos logísticos (apenas 2,59 % do total delas). A classe dos Nitossolos, apesar de ter

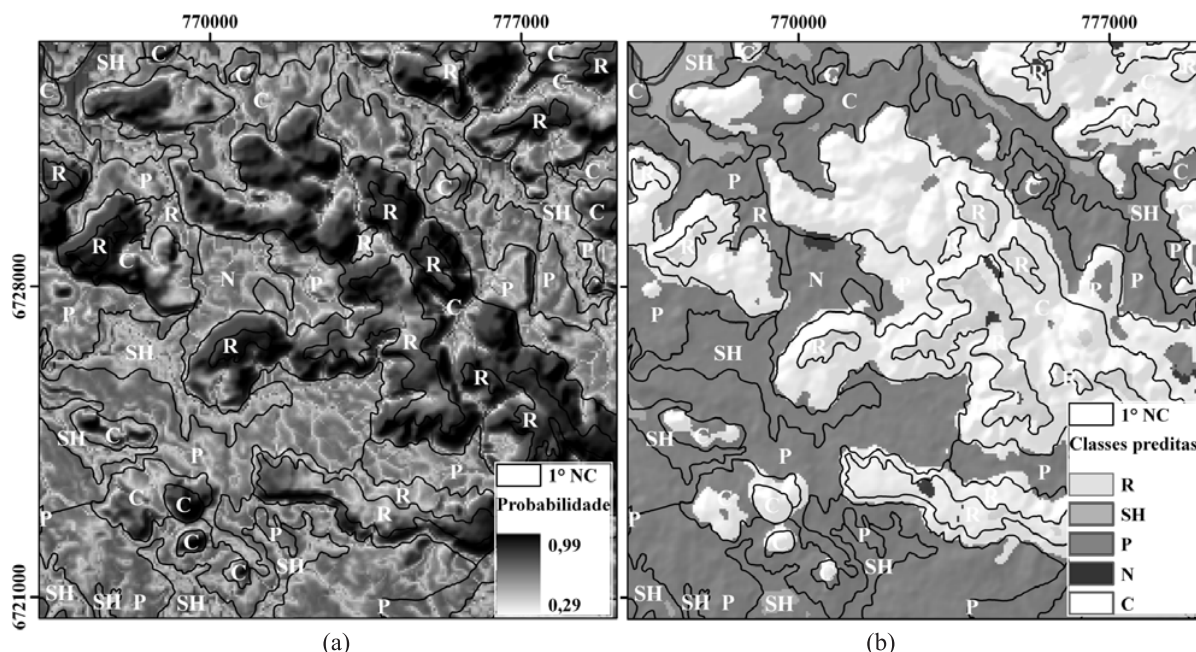


Figura 3. Estrato da área total utilizada no estudo. (a) agrupamento em um único PI dos maiores valores de probabilidade de cada classe de solo. (b) mapa de classes preditas no 1º Nível Categórico (1º NC) na mesma região da imagem esquerda. Sobreposto a essas imagens está o mapa original usado para treinamento. Coordenadas E e N em metros, UTM/SAD69, Fuso 21. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).

sido predita, reproduz visualmente muito pouco do que estava localizado no mapa original de solos, pois estes são também pouco representativos na área total do estudo; consequentemente, pequeno número de amostras foi utilizado para treinar os modelos a partir desta classe (apenas 0,87 %).

Proporção relativa das amostras e posição na paisagem

A relação entre a proporção relativa das classes nas 70.000 amostras, a posição na paisagem e o erro de espacialização de cada uma das classes de solo em ambos os níveis categóricos pode ser visualizada na figura 4.

A figura 4a apresenta as classe preditas considerando apenas o 1º NC do SiBCS. Para a classe dos Neossolos (primeira barra à esquerda) do mapa original, observa-se que outras classes, além dos Neossolos, foram preditas, principalmente a classe dos Cambissolos, pelo fato de esses solos ocuparem praticamente a mesma posição da paisagem. Em menor proporção também foram espacializados Argissolos e Nitossolos.

Nota-se que mais de 75 % dos solos hidromórficos foram preditos corretamente; o restante foi erroneamente predito como Argissolos. Esse confundimento ocorreu em razão de a classe dos Argissolos estar numa posição da paisagem muito próxima à ocupada pelos solos hidromórficos.

A classe dos Argissolos foi predita corretamente, em relação ao mapa original, em aproximadamente 85 %, sendo o restante predito como solos hidromórficos e Cambissolos. Essa elevada acurácia de predição foi, provavelmente, devido à grande proporção relativa da classe dos Argissolos no total das amostras utilizadas para o treinamento dos modelos (Figura 4b), uma vez que é a classe predominante na área de estudo.

Tanto a classe dos Nitossolos quanto a dos Plintossolos não foram preditas adequadamente, possivelmente, devido à pequena proporção relativa delas no total das amostras utilizadas (0,87 e 2,59 % respectivamente – Figura 4b). Essa pequena proporção relativa deve-se à baixa representatividade dessas classes no mapa original. Foi observado que uma proporção relativa baixa dificulta a geração dos modelos, o que está de acordo com Bailey et al. (2003), no sentido de que, para melhorar a capacidade dos modelos em prever as classes de solo, desconsideram-se as classes com área menor do que 5 % da área total do mapa original.

Buscando ampliar a diversidade de classes para testar os modelos, foi utilizado o 2º NC do SiBCS (Figura 4c). Observa-se que a classe dos Argissolos (1º NC), quando subdividida em Argissolos Vermelhos, Argissolos Vermelho-Amarelos, Argissolos Bruno-Acinzentados e Argissolos Acinzentados (2º NC), apresentou proporção relativa de 3,48; 22,40; 13,35; e 7,61 %, respectivamente (Figura 4d).

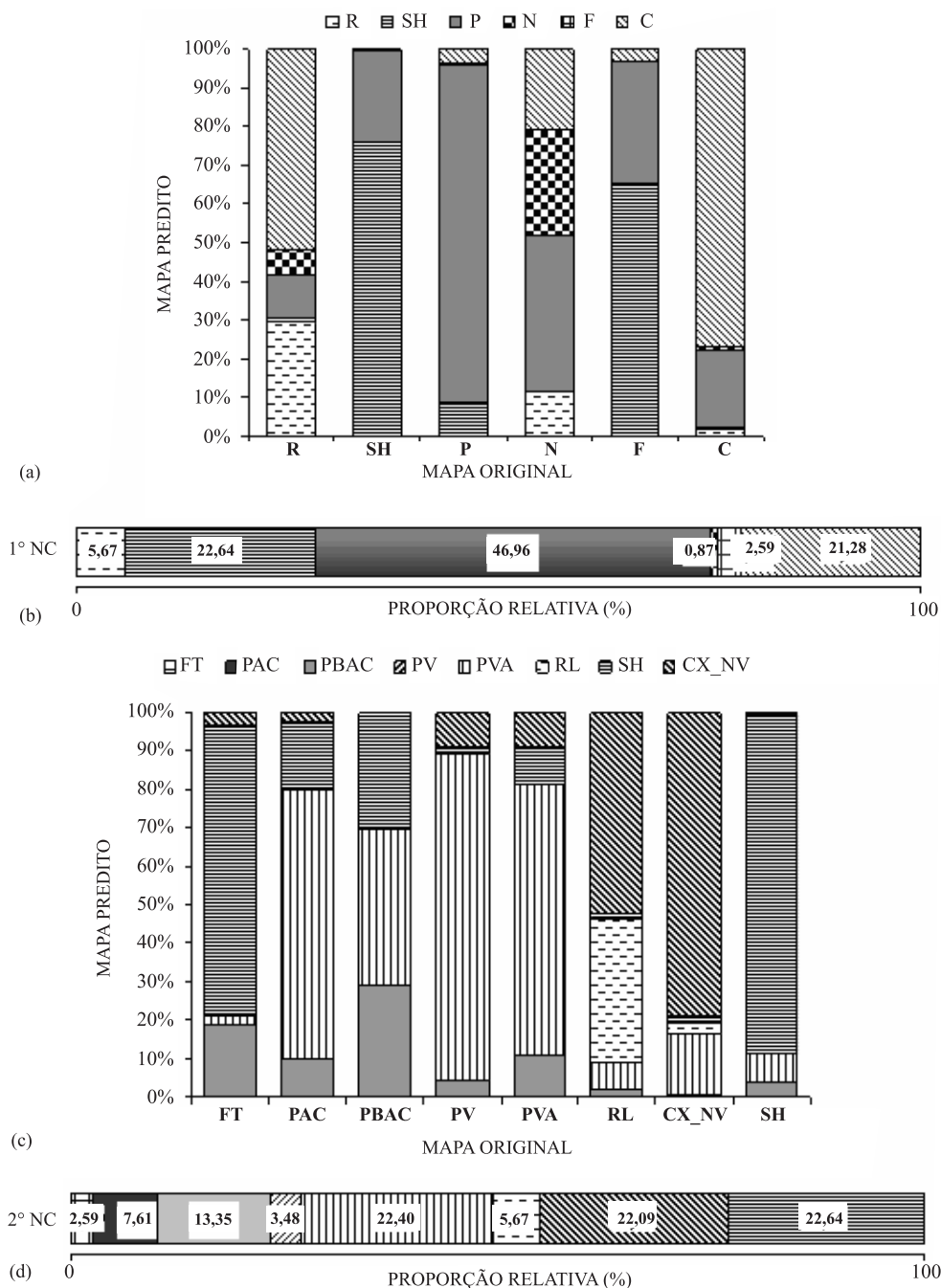


Figura 4. Gráficos de barras verticais: valores oriundos da matriz de erros entre o mapa original e o mapa predito no 1º Nível Categórico (a) e no 2º Nível Categórico (c). Gráficos de barras horizontais: proporção relativa de cada classe de solos no conjunto total de 70.000 amostras utilizadas para gerar os modelos no 1º Nível Categórico (b) e no 2º Nível Categórico (d). Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F), associação Cambissolo Háplico e Nitossolo Vermelho (CX_NV), Argissolo Vermelho (PV), Solos Hidromórficos (SH), Neossolo Litólico (RL), Argissolo Vermelho-Amarelo (PVA), Argissolo Acinzentado (PAC), Argissolo Bruno-Acinzentado (PBAC) e Plintossolo Argilúvico (FT). Nível Categórico (NC).

As classes dos Plintossolos Argilúvicos (primeira coluna), dos Argissolos Acinzentados (segunda coluna) e dos Argissolos Vermelhos (quarta coluna), com proporção relativa de 2,59, 7,61 e 3,48 %, respectivamente, foram selecionadas para a análise de variabilidade genética.

respectivamente, não foram espacializadas sob a paisagem (Figura 4c), corroborando o que havia sido encontrado quando do teste dos modelos para o 1° NC do SiBCS. Esses dados confirmam que classes pouco representativas nas áreas onde está sendo aplicada a RLM poderão não ser preditas pelos modelos.

Quanto às classes dos Argissolos Bruno-Acinzentados e Argissolos Vermelho-Amarelos, com proporção relativa de 13,35 e 22,40 %, respectivamente, pode-se verificar que as áreas adequadamente preditas são proporcionais as amostras de cada uma das classes. Os Argissolos Bruno-Acinzentados foram adequadamente espacializados em 30 % dos locais, de acordo com o mapa original (terceira coluna, Figura 4c), e os Argissolos Vermelho-Amarelos alcançaram índice de acerto próximo a 70 % em relação ao mapa original (quinta coluna, Figura 4c).

Valores igualmente elevados de acerto, do mapa predito em relação ao original, foram alcançados pelas associações Cambissolo Háplico e Nitossolo Vermelho e solos hidromórficos (Figura 4c). Isso deve-se à proporção relativa de ambas as classes de solo – acima de 20 % do total das 70.000 amostras utilizadas no estudo (Figura 4d).

Na figura 4c, percebe-se que o erro dos modelos na predição da adequada classe de solo deu-se novamente entre classes próximas na paisagem. Nas colunas 2, 3, 4 e 5 (quatro classes de Argissolos no 2° NC), os maiores percentuais de confundimento da predição ocorreram entre essas classes. Uma vez que essas classes de solo ocupam posições semelhantes na paisagem, a diluição da proporção relativa do 1° para o 2° NC, sem agregar aos modelos novas covariáveis predictoras, utilizando-se apenas do relevo, a predição adequada torna-se uma tarefa difícil para os modelos RLM. Como a diferenciação entre as classes dos Argissolos no 2° NC é função principalmente da cor, e esta por sua vez pode estar relacionada ao material de origem e, ou, à umidade, sugere-se implementar os modelos levando em consideração esses parâmetros para melhorar a capacidade de predição dessas classes.

Para as demais classes de solo no 2° NC essa situação se repete; o confundimento deu-se sempre entre classes próximas na paisagem. Disso infere-se que, quando do ajuste dos modelos RLM utilizando o mapa existente, estes tiveram dificuldades em extrair a relação solo-paisagem observada no mapa original. Essa dificuldade por parte dos modelos pode ter origem no próprio delineamento que serviu de treinamento, uma vez que o solo não tem uma transição abrupta, como as classes de solo no mapa original (com polígonos cloropléticos), ou, ainda, devido a diferenças muito tênues entre os atributos do terreno (covariáveis ambientais), os quais podem não apresentar nenhum tipo de gradiente na borda dos polígonos das classes de solo.

Boruvka & Penizek (2007) utilizaram redes neurais para a predição de classes de solo e verificaram que classes muito semelhantes sob o ponto de vista

dos processos de formação tendem a ser confundidas pelos modelos. Esses autores observaram que a semelhança entre solos como os *Luvisols* e *Albeluvisols* (IUSS Working Group WRB, 2006) ocasionou incremento de solos incorretamente classificados. O emprego de qualquer método, segundo esses autores, deverá considerar o nível categórico a ser predito em função da heterogeneidade local, além da disponibilidade de informações para a geração dos modelos, como número de perfis para treinamento dos modelos ou número de preditores.

No que se refere à proporção relativa das amostras (Figura 4b,d), pode-se afirmar que as regressões logísticas geram os valores de probabilidades com base não somente nos valores dos preditores, mas também na proporção relativa entre as classes das variáveis-resposta, concordando com observações de Real et al. (2006). Trabalhos de Hengl et al. (2007) e Kempen et al. (2009) mostraram que os modelos logísticos foram dependentes de forte correlação entre os preditores e as classes de solo; para isso, a representatividade mínima de cada classe de solo nos dados utilizados para o treinamento dos modelos precisa ser verificada. Tendo em vista essas informações, pode-se dizer que as classes de solos a serem estudadas devem estar mapeadas em uma proporção relativa mínima, a fim de que os dados submetidos à RLM não apresentem problemas de tendência dos resultados em favor das classes proporcionalmente mais representativas.

CONCLUSÕES

1. As Regressões Logísticas Múltiplas possibilitam gerar mapas de probabilidade de classes de solo na paisagem. A utilização de covariáveis predictoras ligadas apenas a um fator de formação do solo atribui os maiores valores de probabilidade às classes de solo que têm sua gênese ligada a esse fator de formação – no caso deste estudo, Cambissolos e Neossolos relacionados a terrenos forte ondulados e montanhosos.

2. A técnica de modelos lineares generalizados mostrou-se sensível à proporção relativa das amostras utilizadas para gerar os modelos, indicando que aplicações sistemáticas de RLM ao MDS deverão considerar o agrupamento (associação ou inclusão) das classes de solo pouco representativas.

3. Em decorrência de a maior proporção de erros na predição ter ocorrido entre classes próximas na paisagem, a aplicação dessa técnica preditiva deve ser realizada a partir de um maior número possível de covariáveis ambientais representativas, o que possibilitará aos modelos maior poder de distinção entre as diferentes classes na paisagem.

4. Estudos nessa linha de pesquisa deverão propor qual o limite mínimo (*threshold*) de proporção relativa entre as classes de solo para que elas possam ser preditas adequadamente.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de Mestrado concedida ao primeiro autor; e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento e Bolsa PQ concedida ao segundo autor.

LITERATURA CITADA

- BAILEY, N.; CLEMENTS, T.; LEE, J.T. & THOMPSON, S. Modelling soil series data to facilitate targeted habitat restoration: A polytomous logistic regression approach. *J. Environ. Manag.*, 67:395-407, 2003.
- BORUVKA, L. & PENIZECK, V. A test of an artificial neural network allocation procedure using the Czech Soil Survey of Agricultural Land data. In: LAGACHERIE, P.; MCBRATNEY, A. & VOLTZ, M., eds. *Digital soil mapping: an introductory perspective*. Amsterdam, Elsevier, 2007. p.415-424.
- BUI, E.N.; HENDERSON, B.L. & VIERGEVER, K. Knowledge discovery from model of soil properties developed through data mining. *Ecol. Model.*, 191:431-446, 2006.
- CHATTERJEE, S. & HADI, A.S. *Regression analysis by example*. 4.ed. New York, John Wiley & Sons, 2006. 375p.
- CONGALTON, R.G. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens. Environ.*, 37:35-46, 1991.
- DEBELLA-GILO, M. & ETZELMÜLLER, B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena*, 77:8-18, 2009.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. *Sistema brasileiro de classificação de solos*. 2.ed. Rio de Janeiro, Embrapa Solos, 2006. 306p.
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE – ESRI. *ESRI, ArcGIS 93*, Redlands. California, Software, 2008.
- FIGUEIREDO, S.R.; GIASSON, E.; TORNQUIST, C.G. & NASCIMENTO, P.C. Uso de regressões logísticas múltiplas para mapeamento digital de solos no planalto médio do RS. *R. Bras. Ci. Solo*, 32:2779-2785, 2008.
- GIASSON, E.; CLARKE, R.T.; INDA JUNIOR, A.V.; MERTEN, G.H. & TORNQUIST, C.G. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. *Sci. Agric.*, 63:262-268, 2006.
- HARTEMINK, A.E. & MCBRATNEY, A.B. A soil science renaissance. *Geoderma*, 148:123-129, 2008.
- HENGL, T.; TOOMANIAN, N.; REUTER, H.I. & MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, 140:417-427, 2007.
- IUSS Working Group WRB. *World reference base for soil resources*. Rome, FAO, 2006. 145p. (World Soil Resources Reports, 103)
- KEMPEN, B.; BRUS, D.J.; HEUVELINK, G.B.M. & STOOORVOGEL, J.J. Updating the 1:50.000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 125:311-326, 2009.
- KLAMT, E.; FLORES, C.A. & CABRAL, D.R. *Solos do Município de São Pedro do Sul*. Santa Maria, CCR/UFSM, 2001. 96p.
- LAGACHERIE, P. & MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. In: LAGACHERIE, P.; MCBRATNEY, A. & VOLTZ, M., eds. *Digital soil mapping: An introductory perspective*. Amsterdam, Elsevier, 2007. p.3-22.
- MCBRATNEY, A.B.; MENDONÇA SANTOS, M.L. & MINASNY, B. On digital soil mapping. *Geoderma*, 117:3-52, 2003.
- MERMUT, A.R. & ESWARAN, H. Some major developments in soil science since the mid-1960s. *Geoderma*, 100:403-426, 2001.
- OHLMACHER, G.C. & DAVIS, J.C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng. Geol.*, 69:331-343, 2003.
- PENG, C.-Y.J.; SO, T.-S.H.; STAGE, F.K. & JOHN, E.P. St. The Use and Interpretation of Logistic Regression in Higher Education Journals: 1988–1999. *Res. Higher Educ.*, 43:259-293, 2002.
- REAL, R.; BARBOSA, A. & VARGAS, J. Obtaining environmental favourability functions from logistic regression. *Environ. Ecol. Stat.*, 13:237-245, 2006.
- VOLTZ, M.; LAGACHERIE, P. & LOUCHARTE, X. Predicting soil properties over a region using sample information from a mapped reference area. *Eur. J. Soil Sci.*, 48:19-30, 1997.
- WILSON, J.P. & GALLANT, J.C. Digital terrain analysis. In: WILSON, J.P. & GALLANT, J.C., eds. *Terrain analysis: Principles and applications*. New York, Wiley & Sons, 2000. p.1-27.