

Pedometria aplicada à predição de classes de solos utilizando de regressões logísticas múltiplas.

Alexandre ten Caten¹

Ricardo Simão Diniz Dalmolin²

Luiz Fernando Chimelo Ruiz³

Elódio Sebem⁴

Rudiney Soares Pereira⁵

^{1,2}Universidade Federal de Santa Maria – Departamento de Solos
Av. Roraima, 1000 – CCR II – cep: 97105/900 – Santa Maria / RS

¹acaten@yahoo.com.br, ²dalmolinrsd@gmail.com

^{3,4}Colégio Politécnico da UFSM / Av. Roraima, 1000 cep: 97105/900 – Santa Maria / RS

³ruiz.ch@gmail.com, ⁴elodiosebem@gmail.com

⁵Universidade Federal de Santa Maria – Departamento de Engenharia Rural

Av. Roraima, 1000 – CCR II – cep: 97105/900 – Santa Maria / RS

⁵rudiney.s.pereira@gmail.com

Abstract: Pedometry can be defined as the numerical development of statistical models of the relationship between environmental variables and soil which are then applied to a geographic database to create a predictive map. The objective of this study was to use a soil survey on a reference area to train multiple logistic regressions using terrain attributes as variables for predicting of soil spatial distribution. Nine terrain attributes derived from a digital elevation model were used as explanatory covariates, which were: elevation, slope, distance to nearest stream, planar curvature, profile curvature, illumination, natural logarithm of the contributing area, topographic wetness index and sediment transport capacity. To train the logistic models soil surveys were conducted by photointerpretation and field checking of two reference areas in the municipality of São Pedro do Sul / RS. The kappa index of the predicted map was 51.15%, soil units with the smallest area on the reference soil map were not spatialized. The kappa index of the predicted map checked with field data was 37.9%, the error matrix shows that the errors of prediction were more common among classes close to each other. The use of multiple logistic regressions spatialized soil classes from a reference area with quality compatible with a range of applications, the methodology is an alternative for production of soil maps in the continental Brazilian territory.

Palavras-chave: Mapeamento digital de solos, índice de umidade topográfica, capacidade de transporte de sedimento, área de referência, SRTM, digital soil mapping, topographic wetness index, sediment transport capacity, reference area, SRTM.

1. Introdução

1.1 Mapeamento Digital de Solos

As demandas da sociedade pela informação solo têm se modificado, fazendo emergir novas áreas de interesse como qualidade do solo e do ambiente, reconhecimento de processos de degradação e arenização e enfoques em poluição do solo. Estas demandas têm dado aos levantamentos de solos um foco maior em modelagem quantitativa com o acompanhamento de questões envolvendo acurácia e incertezas (MCBRATNEY et al., 2000). Estudos quantitativos em solos têm sido desenvolvidos na área da ciência do solo denominada Pedometria. Podendo ser definida como o desenvolvimento de modelos numéricos ou estatísticos das relações entre variáveis ambientais e o solo, os quais são então aplicados a um banco de dados geográfico para criar um mapa preditivo (SCULL et al., 2003).

McBratney et al. (2003) propuseram uma nova abordagem para a equação dos “Fatores de Formação do Solo” de Jenny (1941). Segundo os autores classes de solo podem ser preditas utilizando-se:

$$Sc = f(s, c, o, r, p, a, n)$$

Onde, Sc = classe do solo, s = próprio solo, c = clima, o = organismos, r = relevo, p = material de origem, a = tempo, n = coordenadas geográficas. Segundo os autores após a equação “f” ser ajustada em um conjunto de dados, esta poderá ser usada em outros locais onde somente as variáveis ambientais são conhecidas. Neste estudo “f” estará sendo operacionalizado pelas Regressões Logísticas Múltiplas (RLM).

1.2 Regressões logísticas múltiplas

Em casos onde o resultado de uma inferência pode ser dado sob muitas categorias (politômica), uma alternativa é trabalhar com a probabilidade de ocorrência de cada uma das categorias, para tanto se aplica a RLM. Uma vez que a RLM não tem qualquer exigência para sua aplicação a respeito da distribuição das variáveis explicativas, não necessitando 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 (CHATTERJEE e HADI, 2006).

A função logarítmica ou logit é a razão entre a probabilidade do evento ocorrer e a probabilidade deste 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 através 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 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.

1.3 Tecnologias envolvidas

O desenvolvimento da análise digital do terreno e tecnologias auxiliares tem criado uma oportunidade para novos métodos científicos serem aplicados nos levantamentos de solos. A análise do terreno tem o potencial de melhorar os levantamentos pedológicos em três áreas. Podendo ser utilizado para: gerar informações ambientais em alta resolução para direcionar o uso da terra (radiação, declividade, etc.), criar estratificações ambientais explícitas, e fornecer preditores espaciais quantitativos para classes de solo (MCKENZIE et al., 2000). Para Moore et al. (1993) em muitos locais, o desenvolvimento do solo em uma catena ocorre em resposta ao movimento da água através e sob a paisagem. Atributos do terreno podem caracterizar os fluxos e, desta forma, os atributos do solo. O terreno modifica a distribuição de processos hidrológicos, erosivos e a temperatura do solo. Desta forma afeta a distribuição do intemperismo mineral, lixiviação, erosão, sedimentação, decomposição, desenvolvimento do perfil, e, finalmente, os atributos do solo. Programas para a geração e espacialização de preditores têm sido desenvolvidos em estudos relacionados a análise do terreno, um exemplo é o *Terrain Analysis Programs for the Environmental Sciences* (TAPES) (WILSON e GALLANT, 2000). Este software tem sido aplicado com sucesso em mapeamento de solos, os quais relacionam a distribuição espacial do solo a parâmetros do terreno (MCKENZIE et al., 2000).

O objetivo do estudo foi utilizar classes de solos em um mapa produzido em área de referência como variáveis resposta, e atributos de terreno como variáveis explicativas em modelos logísticos múltiplos para a predição de classes de solos em áreas não mapeadas.

2. Metodologia de Trabalho

2.1 Caracterização da área de estudo

A área de estudo localiza-se no município de São Pedro do Sul, situado na região central do estado do RS. No município foram realizados levantamentos pedológicos baseados na hipótese de Área de Referência (LAGACHERIE et al., 1995), a qual assume que é possível delimitar pequenas regiões naturais, as quais irão conter um finito número de classes de solos, ocorrendo em associações com outras classes, e formando um padrão reconhecível e repetidamente distribuído. Em duas áreas do município foram executados levantamentos convencionais de solos as quais foram utilizadas como áreas de referência para treinar os modelos preditivos. Visando a verificação e checagem das classes de solos das áreas de referência, foram amostrados 56 pontos no campo com descrição de perfis e tradagens. As classes de solos foram descritas até o nível de subordem. As informações de campo também foram utilizadas para testar a acurácia do mapa gerado.

2.2 Atributos do terreno

A partir de dados de um Modelo Digital de Elevação oriundo do *Shuttle Radar Topography Mission* (SRTM) interpolado para uma resolução de 50 m foram derivados os atributos de terreno: ELEV (Elevação), DECL (Declividade), DIST (Distância à Drenagem), CPLN (Curvatura Planar), CPRF (Curvatura de Perfil), ILUM (Radiação relativa disponível), LNAC (Logaritmo natural da área de contribuição), IUT (Índice de umidade topográfica) e CTS (Capacidade de transporte de sedimento). Inicialmente foi gerada a Área de Contribuição (AC) acima do pixel utilizando o aplicativo TAPES, com o algoritmo FD8 sendo executado em ambiente ArcGIS 9.2 (WILSON e GALLANT, 2000). O valor da AC representa a área acima de determinado ponto da paisagem (pixel) que está contribuindo para o fluxo de água que passa por este ponto. Foi extraído o logaritmo natural do valor da AC, para reduzir a amplitude desta variável, passando a mesma a ser denominada de LNAC neste estudo.

O IUT também denominado de índice topográfico, índice de umidade, e ainda, índice topográfico composto, foi obtido pelo logaritmo natural da razão entre a Área de Contribuição Específica (AC_E) e a tangente da declividade do terreno, equação 2, (WILSON e GALLANT, 2000).

$$IUT = \ln\left(\frac{AC_E}{tg(\beta)}\right) \quad (2)$$

Onde, AC_E foi obtida pela razão entre a AC e a resolução do pixel, equação 3.

$$AC_E = \frac{AC}{resolução_pixel} \quad (3)$$

Os cálculos de IUT e AC_E foram efetuados no aplicativo Análise Espacial (ArcGIS 9.2) pelo procedimento Calculadora Raster, o qual possibilita realizar operações com Plano de Informação (PI) do tipo raster.

Para o cálculo da CTS foi utilizada a equação 4, (WILSON e GALLANT, 2000):

$$CTS = \left(\frac{AC_E}{22.13}\right)^{0.6} \left(\frac{sen(\beta)}{0.0896}\right)^{1.3} \quad (4)$$

Onde, $sen(\beta)$ representa o seno da declividade do terreno. A função foi processada na função Calculadora Raster (ArcGIS 9.2).

A variável representando a ELEV foi obtida diretamente do MDE/SRTM. Tendo sido realizado o procedimento *FillSinks* (ArcGIS 9.2), para correção de depressões espúrias. As informações de DECL, CPLN, CPRF foram derivadas do MDE/SRTM utilizando-se do aplicativo TAPES (WILSON e GALLANT, 2000). A variável ILUM foi obtida através da função *Hillshade* do programa ArcGIS. Com um azimute de 0° e com inclinação solar de 60°.

Para determinação da DIST, foi utilizada a função *Cost Weighted* no programa ArcGIS, onde o peso foi definido como sendo a elevação do terreno.

2.3 Mapa predito a partir do mapa de solos das áreas de referência.

As RLM foram geradas utilizando-se dos dados dos atributos de terreno como variáveis explicativas, e das classes de solos no mapa referência ao nível de subordem (2º Nível Categórico) como variáveis respostas. Para os modelos RLM também só foram consideradas os parâmetros significativos ao nível de 5%. Cada função logit gerou um mapa de probabilidade da existência de determinada classe de solos na paisagem. Estes valores foram reunidos em um único PI, como o maior valor entre os planos definindo a classe predita naquele pixel. O teste da acurácia (índice kappa) dos mapas preditos foi realizado por 1.000 pontos gerados aleatoriamente em cada uma das áreas de referência, bem como, utilizando-se de 56 pontos de campo.

3. Resultados e Discussão

3.1 Mapa predito

Houve parâmetros dos modelos logísticos múltiplos não significativos (Tabela 1). No caso de CTS em nenhuma das classes este parâmetro foi utilizado como parâmetro das funções logit.

Tabela 1 – Valores dos interceptos e dos coeficientes dos atributos do terreno nos modelos logísticos das áreas de referência.

	CX ¹				
	SH	PBAC	PV	RL	PVA
Intercepto	36,384	26,510	17,125	-0,278	20,460
ELEV	-0,077	-0,053	-0,027	0,014	-0,045
DECL	-0,238	-0,031	-0,027	-0,050	-0,198
DIST	-0,001	-0,001	0,0003	0,001	0,0005
CPLN	1,293	-0,070	-2,605	-2,018	-1,806
CPRF	4,221	4,147	3,331	1,554	3,000
ILUM	-0,090	-0,053	-0,023	0,001	-0,026
LNAC	-1,572	-1,472	-1,383	-0,311	-1,421
IUT	1,096	1,077	0,989	0,084	0,835
CTS	0,007	-0,006	-0,009	0,010	0,016

¹Classe de referência para os modelos logísticos. Valores em negrito não são significativos em nível de 5%. 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 (*Likelihood Ratio Test*). Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno Acinzentado (PBAC), Solo Hidromórfico (SH), Elevação (ELEV), Declividade (DECL), Curvatura Planar (CPLN), Curvatura de Perfil (CPRF), Distância à Drenagem (DIST), Iluminação (ILUM), Área de Contribuição (LNAC), Índice de Umidade Topográfica (IUT) e Capacidade de Transporte de Sedimento (CTS).

A utilização dos atributos do terreno como variáveis explicativas possibilitou a espacialização dos Solos Hidromórficos (Associação Gleissolos Melânicos e Plintossolos Argilúvicos), estes solos foram posicionados ao longo das drenagens dos rios. No entanto, em grande parte das regiões de várzea onde esta classe estava mapeada no mapa de referência, foi espacializada no mapa predito a classe dos Argissolos Bruno Acinzentados.

A posição dos Neossolos na paisagem corresponde visualmente ao que estava delineado no mapa referência. A conferência entre os Argissolos Vermelhos em ambos os mapas demonstra que esta classe foi visualmente reproduzida (Figura 1).

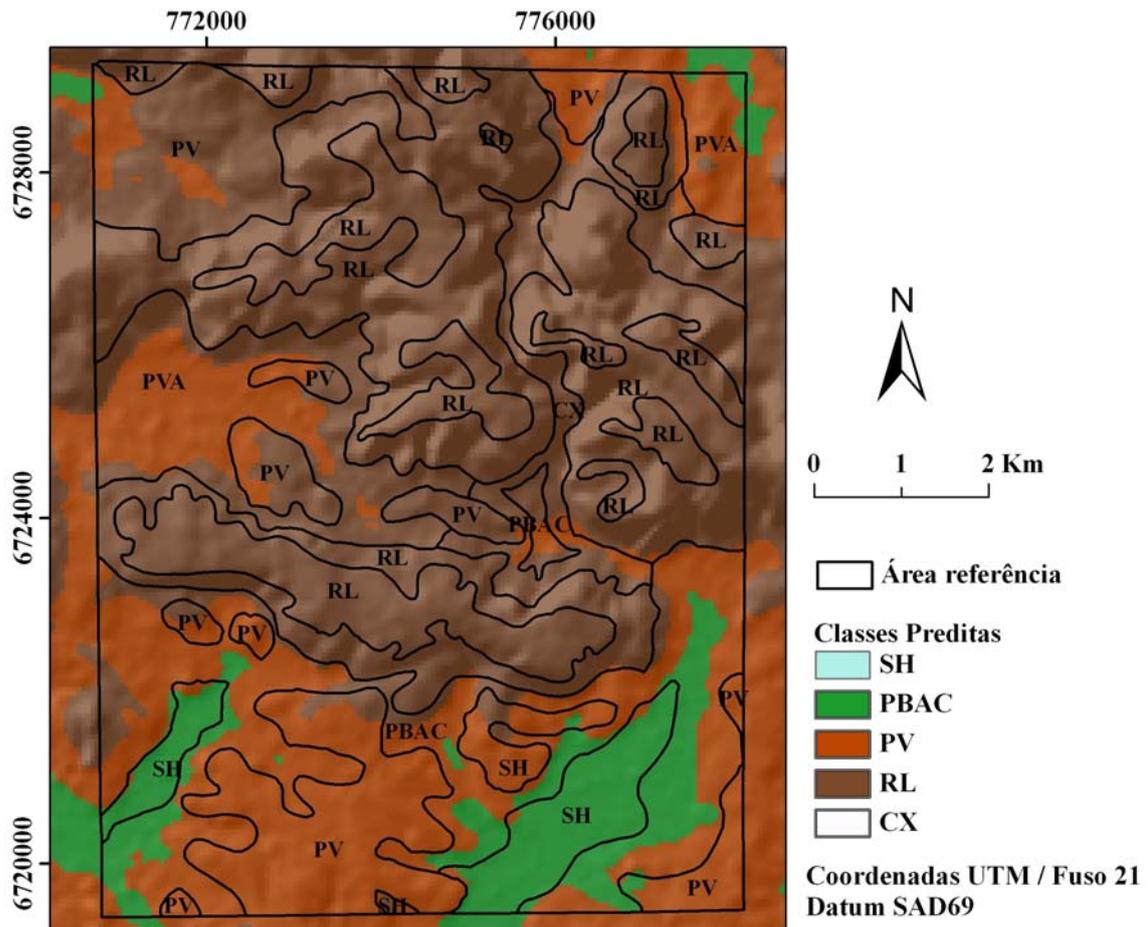


Figura 1 – Sessão do mapa de solos predito com os delineamentos de uma das áreas de referência sobreposta. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno Acinzentado (PBAC) e Solo Hidromórfico (SH).

O mapa predito reproduziu em torno de 50% as informações que constavam no mapa referência (índice kappa) (Tabela 2). A acurácia do mapeador (AM) para os Solos Hidromórficos demonstra que estes foram confundidos com os Argissolos Bruno Acinzentados em grande parte dos locais amostrados para o teste da acurácia. As classes Argissolo Vermelho-amarelo e Cambissolo Háplico não foram espacializadas pelo modelo preditivo, estas duas classes corresponderam as menores áreas mapeadas nas áreas de referência, 7,92 e 3,52 ha, respectivamente. Para melhorar a capacidade dos modelos logísticos em prever os solos ao nível de série Bailey et al. (2003) desconsideraram as classes com área menor do que 5% da área total. Real et al. (2006) afirmam que as regressões logísticas geram os valores de probabilidades baseadas não somente nos valores dos preditores, mas também na proporção relativa entre as classes das variáveis resposta. Pode-se assim supor que estas classes não tenham sido adequadamente espacializadas devido a RLM ser sensível a proporção relativa entre as classes preditas.

Tabela 2 – Matriz de erros entre o mapa da área de referência e o predito produzido a partir dos atributos do terreno como variáveis dos modelos logísticos. Acurácia do mapeador (AM), acurácia do usuário (AU) e acurácia geral e índice kappa (K).

Mapa predito	Mapa referência						Total	AU(%)
	SH	PBAC	PV	RL	PVA	CX		
SH	0,46	0,30	0,00	0,00	0,00	0,00	0,76	60,53
PBAC	6,02	13,61	7,29	0,00	1,42	0,00	28,34	48,02
PV	1,01	8,60	26,32	0,30	4,35	0,40	40,99	64,21
RL	0,15	0,76	5,26	21,41	0,00	2,33	29,91	71,58
PVA	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
CX	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Total	7,64	23,28	38,87	21,71	5,77	2,73	100,00	
AM(%)	6,02	58,46	67,71	98,62	0,00	0,00		
AG(%)	61,79		K¹(%)	51,15				

erro padrão assintótico = $\pm 0,013$ e intervalo de confiança 95% = 0,48...0,53. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno Acinzentado (PBAC) e Solo Hidromórfico (SH).

A tabela acima confirma pela acurácia do mapeador que os Neossolos Litólicos foram adequadamente posicionados em relação ao mapa referência (98,62%), embora em torno de 30 % dos locais esta classe não condiz com o que foi predito (acurácia do usuário de 71,58%). Uma verificação de como as classes de solos foram confundidas pode ser feita analisando a figura 2. Nos Solos Hidromórficos verifica-se que os mesmos foram confundidos com os Argissolos Bruno Acinzentados em grande parte dos locais. Enquanto que os Neossolos Litólicos apresenta-se adequadamente preditos. A figura 2 demonstra que as classes de solos foram confundidas com classes espacialmente próximas na paisagem. Nas regiões de várzea onde efetivamente encontravam-se os Solos Hidromórficos foram espacializados os Argissolos Bruno Acinzentados, nas regiões de coxilha as classes dos Argissolos confundiram-se entre elas na medida em que se vai dos locais mais baixos para os mais altos da paisagem.

Em relação a verdade de campo os Solos Hidromórficos foram adequadamente espacializados. No entanto, é importante frisar que esta é a realidade do conjunto de 56 pontos visitados. Não sendo este número nem muito elevado, e nem correspondente a pontos fora dos locais utilizados como mapa referência.

O índice K do mapa predito checado a campo foi de 37,9%, similar os valores encontrados por Hengl et al. (2007), que utilizando de RLM espacializaram classes de solos para o território do Iran com uma acurácia kappa de 36,7%. Para estes autores os 5.200 perfis de solo utilizados para treinar os modelos preditivos, não estavam adequadamente distribuídos pelo território, além do solo ser de difícil predição devido a sua grande variabilidade.

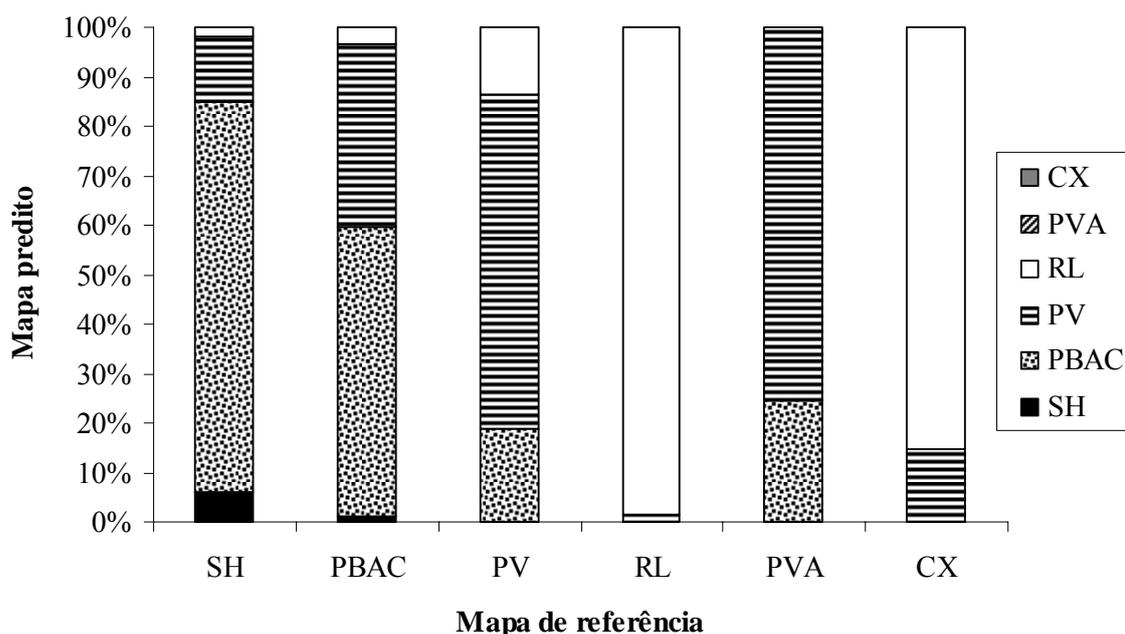


Figura 2 - Proporção relativa dos erros de inferência entre o mapa referência e o mapa predito. Cambissolo Háplico (CX), Argissolo Vermelho (PV), Argissolo Vermelho-amarelo (PVA), Neossolo Litólico (RL), Argissolo Bruno Acinzentado (PBAC) e Solo Hidromórfico (SH).

Os erros de predição comparando-se o mapa predito aos pontos de campo revelam um padrão similar ao da comparação para com o mapa referência. As classes foram confundidas pelos modelos predominantemente entre as mais próximas na paisagem. Embora houve confusão também entre os Argissolos Bruno Acinzentados e os Neossolos Litólicos, isto demonstra que para melhorar a capacidade preditiva dos modelos outros preditores teriam de ser utilizados. Mapas temáticos de geologia e uso da terra podem ser utilizados em RLM uma vez que a variável explicativa destes modelos pode ser contínua ou categórica. Uma segunda alternativa para a melhoria da capacidade dos modelos em distinguir entre classes de solos muito próximas na paisagem seria através de um Modelo Digital de Elevação com uma maior resolução espacial. Uma maior resolução permitiria que características mais tênues do relevo pudessem ser corretamente associadas a cada uma das classes de solos. Neste estudo foram utilizados apenas nove parâmetros de terreno em uma resolução de 50 metros, pode estar ocorrendo que gradientes muito suaves da paisagem estejam impossibilitando que os modelos especializem adequadamente as classes de solos.

3.2 Discussão geral do estudo

O aumento da capacidade preditiva dos modelos utilizando-se de RLM passará pela utilização de variáveis explicativas que tenham fortes relações com as classes de solos na região. Outros atributos de terreno, mapas geológicos e de uso da terra, além de uma maior resolução espacial são alternativas para futuros estudos utilizando-se desta técnica.

Outro ponto a ser considerado é a sensibilidade das RLM quanto a proporção relativa das áreas utilizadas para o treinamento dos modelos. Caso ocorram áreas proporcionalmente muito pequenas de algumas das classes de solos, estas deverão ser preditas como inclusões, associações ou complexos de solos.

Embora neste estudo os mapas preditos não tenham sido checados com informações fora das áreas utilizadas para o treinamento dos modelos, verifica-se que esta metodologia tem potencial para a geração da informação solo em áreas não mapeadas. Sendo importante para o emprego da metodologia primeiramente definir a dimensão (raio de abrangência) até a qual a

região pode ser considerada homogênea, nela a área de referência ainda poderá ser considerada válida, mesmo afastada da área onde foi originalmente concebida.

Como a RLM gera um mapa de probabilidade de se encontrar cada uma das classes de solos na paisagem, esta informação precisa ser melhor explorada em futuros estudos. Supõem-se que estes mapas de probabilidade, variando de zero a um, possam ser utilizados como pesos ou ponderadores em modelos que se utilizem das classes de solos.

4. Conclusões

As Regressões Logísticas Múltiplas espacializam as classes de solos na paisagem, embora classes relativamente não representativas nos mapas de treinamento não sejam acuradamente mapeadas pela metodologia.

A inferência da distribuição espacial de classes de solos a partir de atributos de terreno estará limitada a capacidade dos atributos utilizados em representar a complexidade do ambiente de formação dos solos na região.

Agradecimentos

A CAPES pela bolsa de mestrado concedida ao primeiro autor.

Referências Bibliográficas

BAILEY, N.; CLEMENTS, T.; LEE, J.T.; THOMPSON, S. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. **Journal of Environmental Management**, v. 67, n. 4, p. 395-407, 2003.

CHATTERJEE, S.; HADI, A.S. **Regression analysis by example**. John Wiley & Sons, 4. ed., 2006. 375 p.

FIGUEIREDO, S.R. **Mapeamento supervisionado de solos através do uso de regressões logísticas múltiplas e sistema de informações geográficas**. 2006. 96f. Dissertação (Mestrado em Ciência do Solo), Universidade Federal do Rio Grande do Sul, Porto Alegre.

HENGL, T.; TOOMANIAN, N.; REUTER, H.I.; MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, Amsterdam, v. 140, n. 4, p. 417-427, 2007.

LAGACHERIE, P.; LEGROS, J.P.; BURFOUGH, P.A. A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. **Geoderma**, Amsterdam, v. 65, n. 3-4, p. 283-301, 1995.

MCBRATNEY, A.; MENDONÇA SANTOS, M.L.; MINASNY, B. On digital soil mapping. **Geoderma**, Amsterdam, v. 117, n. 1-2, p. 3-52, 2003.

MCBRATNEY, A.; ODEH, I.O.A.; BISHOP, T.F.A.; DUNBAR, M.S.; SHATAR, T.M. An overview of pedometric techniques for use in soil survey. **Geoderma**, Amsterdam, v. 97, n. 3-4, p. 293-327, 2000.

MCKENZIE, N.J.; GESSLER, P.E.; RYAN, P.J.; O'CONNELL, D.A. **The role of terrain analysis in soil mapping**. In: WILSON, J. P.; GALLANT, J. C. (Ed.). **Terrain analysis: principles and applications**. New York: Wiley & Sons, 2000, p. 245-265

REAL, R.; BARBOSA, A.; VARGAS, J. Obtaining Environmental Favourability Functions from Logistic Regression. **Environmental and Ecological Statistics**. v. 13, n. 2, p. 237-245, 2006.

SCULL, P.; FRANKLIN, J.; CHADWICK, O.A.; MCARTHUR, D. Predictive soil mapping: a review. **Progress in Physical Geography**, v. 27, p. 171-197, 2003.

WILSON, J. P.; GALLANT, J. C. **Digital terrain analysis**. In: _____ (Ed.). **Terrain analysis: principles and applications**. New York: Wiley & Sons, 2000, p. 1-27