

## Componentes principais como preditores no mapeamento digital de classes de solos

### Principal components as predictor variables in digital mapping of soil classes

Alexandre ten Caten<sup>I\*</sup> Ricardo Simão Diniz Dalmolin<sup>II</sup> Fabrício de Araújo Pedron<sup>II</sup>  
Maria de Lourdes Mendonça Santos<sup>III</sup>

#### RESUMO

*Tecnologias disponíveis para a observação da Terra oferecem uma grande gama de informações sobre componentes ambientais que, por estarem relacionadas com a formação dos solos, podem ser usadas como variáveis preditoras no Mapeamento Digital de Solos (MDS). No entanto, modelos com um grande número de preditores, bem como a existência de multicolinearidade entre os dados, podem ser ineficazes no mapeamento de classes e propriedades do solo. O objetivo deste estudo foi empregar a Análise de Componentes Principais (ACP) visando a selecionar e diminuir o número de preditores na regressão logística múltipla multinomial (RLMM) utilizada no mapeamento de classes de solos. Nove covariáveis ambientais, ligadas ao fator de formação relevo, foram derivadas de um Modelo Digital de Elevação e denominadas variáveis originais, estas foram submetidas à ACP e transformadas em Componentes Principais (CP). As RLMM foram desenvolvidas utilizando-se atributos de terreno e as CP como variáveis explicativas. O mapa de solos gerado a partir de três CP (65,6% da variância original) obteve um índice kappa de 37,3%, inferior aos 48,5% alcançado pelo mapa de solos gerado a partir de todas as nove variáveis originais.*

**Palavras-chave:** *pedometria, análise estatística multivariada, levantamento de solos.*

#### ABSTRACT

*Available technologies for Earth observation offer a wide range of predictors relevant to Digital Soil Mapping (DSM). However, models with a large number of predictors, as well as, the existence of multicollinearity among the data, may be ineffective in the mapping of classes and soil properties. The aim of this study was to use the Principal Component Analysis*

*(PCA) to reduce the number of predictors in the multinomial logistic regression (MLR) used in soil mapping. Nine environmental covariates, related to the relief factor of soil formation, were derived from a digital elevation model and named the original variables, which were submitted to PCA and transformed into principal components (PC). The MLR were developed using the terrain attributes and the PC as explanatory variables. The soil map generated from three PC (65.6% of the original variance) had a kappa index of 37.3%, lower than the 48.5% achieved by the soil map generated from all nine original variables.*

**Key words:** *pedometric, multivariate statistical analysis, soil survey.*

#### INTRODUÇÃO

Pesquisadores em todo o globo têm buscado propor soluções científicas e tecnológicas dentro do escopo do Mapeamento Digital de Solos (MDS). 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 (SANCHEZ et al., 2009).

Se por um lado o MDS é beneficiado pelo desenvolvimento tecnológico em Sensoriamento Remoto, Sistema de Posicionamento por Satélite e

<sup>I</sup>Instituto Federal Farroupilha (IFF), Campus Júlio de Castilhos, São João do Barro Preto, 98130-000, Júlio de Castilhos, RS, Brasil.  
E-mail: acaten@yahoo.com.br. Autor para correspondência.

<sup>II</sup>Departamento de Solos, Centro de Ciências Rurais (CCR), Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.

<sup>III</sup>Centro Nacional de Pesquisa de Solos, Rio de Janeiro, RJ, Brasil.

Espectroradiometria, por outro, novas dificuldades afloram: é preciso decidir dentro da grande gama de preditores possíveis de serem gerados, por aqueles de maior e melhor potencial preditivo, assim como, conhecer suas inter-relações e se há efeitos de multicolinearidade (HENGL et al., 2007).

Com a intenção de mitigar o efeito das relações lineares entre preditores, HENGL & ROSSITER (2003) empregaram análise de Componentes Principais (CP) para eliminar efeitos de correlação e de diferentes unidades entre os preditores. Segundo esses autores, a aplicação de cada uma das CP como preditores deverá ser feita na medida da variância por ela retida, embora as CP sejam de mais difícil interpretação em relação às covariáveis ambientais originais.

Estudos em MDS trabalham na perspectiva de prever ou extrapolar as informações a respeito da distribuição espacial do solo. Para casos em que o resultado de uma inferência pode ser dado sob muitas categorias ou politômica (classes de solo), uma alternativa é trabalhar com modelos logísticos (TEN CATEN et al, 2009). O modelo de regressão logística permite avaliar a probabilidade de ocorrência de uma variável resposta, considerando os valores de uma série de variáveis independentes que podem ser qualitativas ou quantitativas (HENGL et al., 2007). Por ser utilizada mais que uma variável independente, o modelo é considerado múltiplo e, pelo fato de prever mais que duas classes, é considerado multinomial (ou politômico); logo, o modelo é denominado de regressão logística múltipla multinomial (RLMM), que é uma generalização do modelo logístico simples (COELHO & GIASSON, 2010).

De acordo com CHATTERJEE & HADI (2006), a RLMM não demanda que os dados tenham distribuição normal, homoscedasticidade, correlação linear e observações em mesma escala. As variáveis preditoras podem ser, ainda, uma mistura de dados contínuos, discretos ou binários. Estudos realizados por GIASSON et al. (2006), DEBELLA-GILO & ETZELMÜLLER (2009), KEMPEN et al. (2009), COELHO & GIASSON (2010) e TEN CATEN et al. (2011) demonstram detalhadamente o desenvolvimento desses modelos na predição de classes e propriedades do solo.

O objetivo deste estudo foi empregar a Análise de Componentes Principais (ACP) como alternativa para selecionar e diminuir o número de preditores potencialmente aplicáveis em modelos logísticos múltiplos multinomiais utilizados no mapeamento digital de solos.

## MATERIAL E MÉTODOS

A área de estudo abrange o município de São Pedro do Sul, situado na região central do estado do Rio Grande do Sul, na microrregião de Santa Maria, em uma região transitória entre as regiões fisiográficas do Planalto Médio e Depressão Central. A área territorial do município equivale a 87.400ha, para a qual está disponível um levantamento de solos semidetalhado 1:50.000 (KLAMT et al., 2001).

As nove variáveis preditoras utilizadas no estudo têm relação com a pedogênese local, sendo elas: 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). Todas foram geradas conforme metodologia descrita por TEN CATEN et al. (2009), a partir de um Modelo Digital de Elevação (MDE) com resolução espacial de 50m derivado do *Shuttle Radar Topographic Mission* (SRTM).

A partir dos Planos de Informação (PI) dos nove atributos de terreno, foi extraído um conjunto de 20% de amostras (píxeis) de cada PI que constituíram os dados para a análise estatística multivariada. AACP foi executada no pacote estatístico *Statistical Analysis System* (SAS) (KIM et al., 2007). AACP foi realizada na matriz de correlação dos dados, ou seja, em dados padronizados (média igual a zero e variância igual a um).

Em seguida, os modelos logísticos foram gerados em duas situações, em uma primeira, com os preditores atributos de terreno e, em uma segunda, com os preditores CP. Em todos os modelos logísticos testados, foram considerados apenas os coeficientes significativos a 5% pelo teste de *Wald* CHATTERJEE & HADI (2006).

De posse dos modelos logísticos compostos pelas variáveis preditoras CP e atributos de terreno, foram gerados os PI de probabilidade de ocorrência de cada uma das classes de solos presentes ao nível de Subordem do Sistema Brasileiro de Classificação de Solos – SiBCS (EMBRAPA, 2006). A fase de manipulação dos PI e o geoprocessamento encontra-se pormenorizada em TEN CATEN et al. (2009). O mapa de solos final foi gerado a partir de cada pixel de maior valor de probabilidade em um mesmo ponto da paisagem.

A qualidade do mapa predito foi verificada a partir de um conjunto de pontos de verificação no mapa final. Esses dados foram tabulados em formato ASCII e processados para a geração da matriz de erros,

da qual o Índice Kappa, a Acurácia do Usuário (AU) e a Acurácia do Mapeador (AM) são indicadores utilizados para atestar a qualidade dos mapeamentos preditivos (HENGL et al., 2007; COELHO & GIASSON, 2010).

## RESULTADOS E DISCUSSÃO

### Análise de componentes principais

A análise multivariada possibilitou a transformação dos nove atributos de terreno em um novo conjunto de nove CP, cada qual, em ordem decrescente do primeiro ao nono, retendo parte da variabilidade original dos dados. Nas três primeiras CP, foram retidos dois terços da variância existentes nos atributos do terreno (65,6%). A aplicação dessas três primeiras CP como variáveis preditoras em modelos logísticos representará uma simplificação dos modelos, contudo, como entre a quarta e nona componente está concentrado um terço da variabilidade total dos dados originais (34,4%), a simplificação dos modelos poderá implicar perda de potencial preditivo.

Com até um terço da variância sendo retido nas CP mais elevadas (quarta à nona), é um indicativo de que as variáveis originais possuem um relativo grau de independência. Para HENGL & ROSSITER (2004), que identificaram em torno de 67% da variância nas

três primeiras componentes a partir de atributos de terreno, este resultado não é comum para dados amostrados pela técnica de Sensoriamento Remoto, que usualmente gera dados correlacionados. Segundo esses autores, a maior dificuldade no emprego de ACP está na interpretação das variáveis originais transformadas em CP, as quais não têm significado físico, químico ou biológico.

As três primeiras componentes principais (CP1, CP2 e CP3), as quais retêm dois terços da variabilidade dos dados originais (atributos de terreno), foram utilizadas como variáveis preditoras não correlacionadas em modelos logísticos múltiplos, conforme a seguir.

### Modelo logístico múltiplo multinomial

Os modelos logísticos foram gerados utilizando-se como referência os Solos Hidromórficos (Associação Gleissolos Melânicos + Neossolos Quartzarênicos) por serem estes os solos de maior extensão na área de estudo, embora não existam evidências da influência da classe de referência nos resultados dos modelos logísticos (TEN CATEN et al., 2011).

Valores não significativos para os interceptos e coeficientes dos modelos logísticos não foram utilizados nas respectivas equações (Tabela 1).

Tabela 1 - Valores dos interceptos e coeficientes gerados a partir das covariáveis preditoras e covariáveis resposta em nível de subordem.

Atributos de terreno <sup>1</sup>							
Preditor	FT	PAC	PBAC	PV	PVA	RL	CX_NV
Intercepto	4,742	-8,555	-2,838	-8,173	-7,732	-16,304	-17,080
ELEV	-0,024	0,067	0,035	0,083	0,073	0,114	0,103
DECL	-0,284	0,029	-0,101	-0,062	0,071	0,134	0,264
DIST	0,003	0,002	0,003	0,003	0,003	0,003	0,003
CPLN	-2,134	-3,960	-4,221	-4,887	-4,141	-5,429	-3,850
CPRF	<u>0,763</u>	-0,871	<u>-0,058</u>	-1,963	-1,539	-2,369	-3,230
RADI	-0,026	-0,016	-0,017	-0,033	-0,022	-0,032	-0,020
LNAC	0,403	0,357	0,443	0,386	0,286	<u>-0,030</u>	0,261
IUT	-0,355	-0,214	-0,346	-0,353	-0,160	0,160	0,072
CTS	-0,012	-0,048	-0,039	<u>-0,004</u>	-0,009	-0,006	-0,002
Componentes principais <sup>1</sup>							
Preditor	FT	PAC	PBAC	PV	PVA	RL	CX_NV
Intercepto	<u>-0,384</u>	5,497	4,375	6,379	7,249	7,047	8,258
CP1	-0,013	0,055	0,029	0,080	0,073	0,121	0,122
CP2	-0,097	-0,080	-0,090	-0,083	-0,063	-0,044	-0,038
CP3	<u>0,001</u>	-0,157	-0,097	-0,218	-0,196	-0,304	-0,305

<sup>1</sup>Classe de referência (SH). 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), Plintossolo Argilúvico (FT), 1<sup>o</sup> Componente Principal (CP1), 2<sup>o</sup> Componente Principal (CP2), 3<sup>o</sup> Componente Principal (CP3), 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). Valores sublinhados não foram significativos ao nível de 5% (teste de Wald).

Segundo DEBELLA-GILO & ETZELMÜLLER (2009), preditores não significativos não são correlacionados com as classes de solos a serem preditas e, portanto, devem ser excluídos da fase de implementação dos modelos. Em um estudo realizado na Noruega, esses autores relatam que o atributo de terreno curvatura planar não foi utilizado em nenhum dos modelos por não estar correlacionado com as classes de solos da área de estudo.

A partir dos dados desta tabela, os modelos preditivos foram gerados utilizando os atributos de terreno como preditores, como exemplificado aqui para a classe dos Plintossolo Argilúvico, Equação 1:

$$\log\left(\frac{FT}{SH}\right) = 4,742 - 0,024 * ELEV - 0,284 * \\ DECL + 0,003 * DIST - 2,134 * \\ CPLN - 0,026 * RADI + 0,403 * \\ LNAC - 0,355 * IUT - 0,012 * CTS \quad (1)$$

E utilizando-se das CP como preditores desta classe, Equação 2:

$$\log\left(\frac{FT}{SH}\right) = -0,013 * CP1 - 0,097 * CP2 \quad (2)$$

Verifica-se assim a grande simplificação dos modelos possibilitada pela ACP. Essa situação é desejada principalmente em circunstâncias em que o número de preditores é muito grande, como em aplicações das bandas de sensores remotos hiperespectrais.

#### Mapas de classe de solos

A partir do PI com os maiores valores de probabilidade para todas as classes de solo, geraram-se os mapas de solos para a área do estudo (Figura 1A e 1B). Em uma análise visual destes mapas, é possível afirmar que os modelos foram eficientes em identificar três posições distintas na paisagem: relevo plano (várzeas), relevo suave ondulado a ondulado (coxilhas) e relevo forte ondulado a montanhoso (encosta).

As subordens Associação Cambissolo Háplico + Nitossolo Vermelho e os Neossolos Litólicos foram espacializadas ao longo dos terrenos de encosta; Argissolo Vermelho, Argissolo Vermelho-Amarelo e Argissolo Bruno-Acinzentado foram alocadas em áreas de formação de horizontes profundos e drenados das colinas; e os Solos Hidromórficos ao longo das várzeas em ambientes de má drenagem.

Uma comparação visual entre os mapas da figura 1A e 1B permite verificar que, no segundo caso, as classes foram mapeadas de maneira mais intrincada. Com o uso das CP como preditores, os modelos não contaram com cerca de um terço da variabilidade original

dos dados, o que influenciou no delineamento das classes de solos.

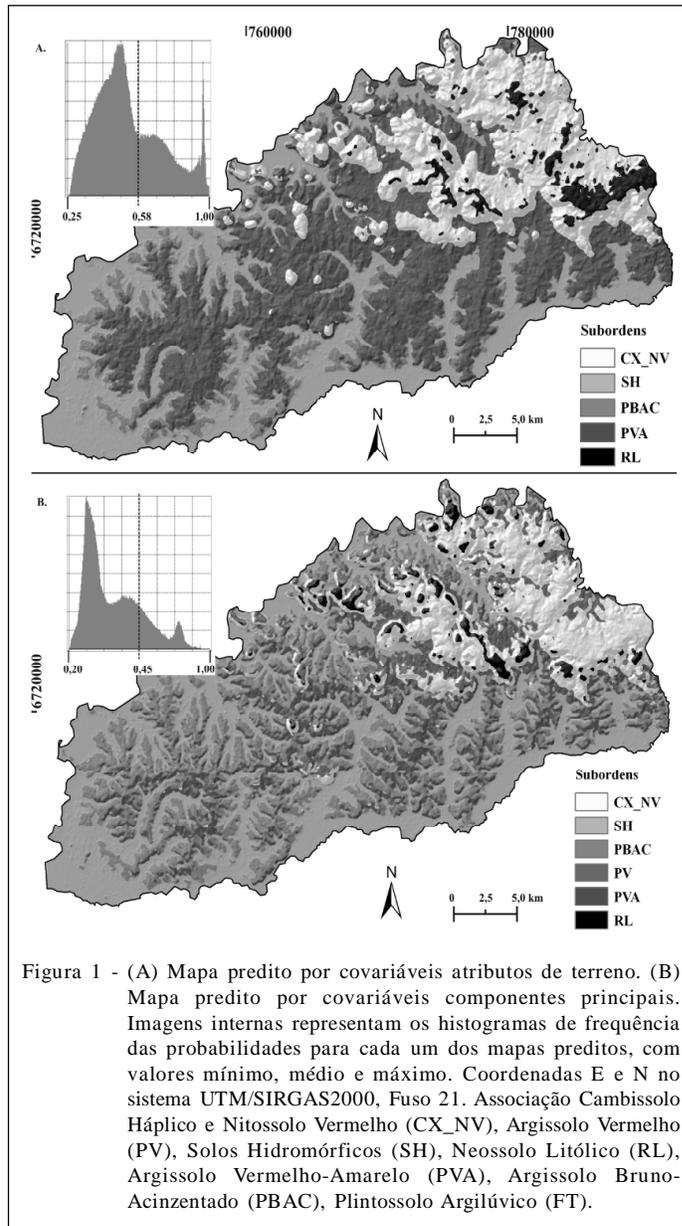
Em ambos os casos, a classe dos Plintossolos Argilúvicos não foi predita e esse fato decorre da pouca representatividade desta classe na área de estudo. Trabalhos de HENGL et al. (2007), KEMPEN et al. (2009) e TEN CATEN (2011) mostraram que os modelos logísticos foram dependentes de uma forte correlação entre os preditores e as classes de solo e, para isso, a representatividade mínima de cada classe de solo nos dados utilizados para o treinamento dos modelos, precisa ser verificada.

Internamente, nas figuras 1A e 1B, estão os histogramas de frequência das probabilidades que deram origem a cada um destes mapas de solos. Verifica-se que, no mapa gerado com a utilização dos atributos de terreno (histograma da Figura 1A), são frequentes valores de probabilidade mais elevados quando comparados com o histograma que utiliza CP (histograma da Figura 1B). Logo, devido ao uso das CP, apenas 65,57% da variabilidade original é utilizada e os mapas gerados possuem uma menor certeza associada, quando comparado aos mapas gerados a partir da totalidade dos dados originais. O menor valor de probabilidade (0,20) e de média (0,45) das probabilidades também ocorreu quando se utilizou das CP como preditores (histograma da Figura 1B).

#### Qualidade dos mapas gerados

Os resultados das matrizes de erro para os mapas gerados encontram-se nas figuras 2A e 2B. O índice kappa dos mapas gerados a partir dos preditores atributos de terreno e CP foi de 48,5%. Este valor é superior aos 36% determinados por GIASSON et al. (2006), aos 38% encontrados por FIGUEIREDO et al. (2008) e de 32,81% em estudos de COELHO & GIASSON (2010). Por sua vez, o uso dos CP como preditores repercutiu em um valor de kappa de 37,3% o que representa uma diminuição de 11 pontos percentuais em relação ao mapa predito utilizando-se de toda a variabilidade dos dados originais. Dessa forma, deve-se ponderar entre o ganho com a simplificação dos modelos gerados e a possibilidade de perda significativa de poder de explicação dos modelos, caso as variáveis originais sejam naturalmente pouco correlacionadas, como ocorreu neste estudo.

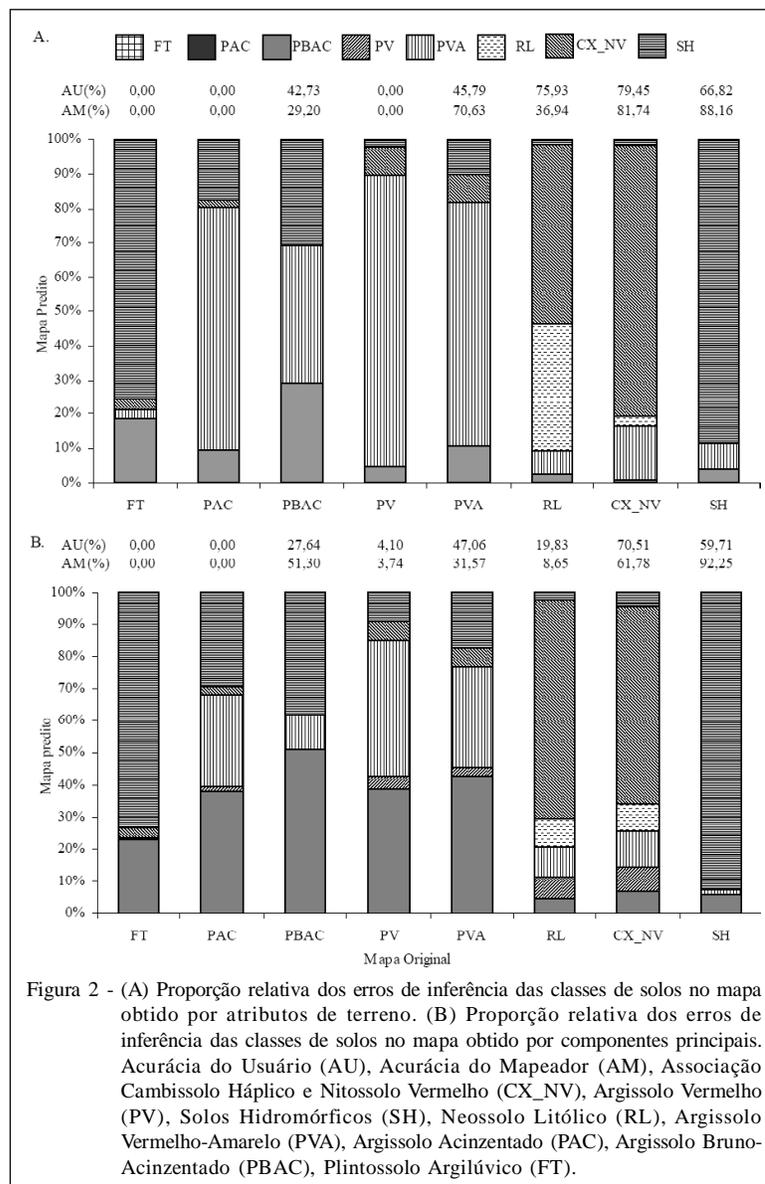
Para HENGL & ROSSITER (2003), o emprego das CP não foi adequado devido ao percentual de variabilidade retida pelas componentes mais elevadas. Esses autores optaram pela não utilização das CP como preditores devido à importância das informações retidas ao longo de todas as componentes para a explicação do fenômeno a ser modelado.



O efeito da aplicação de CP para a AU e AM pode ser visualizado nas figuras 2A e 2B. A análise da AU, a qual está relacionada a erros ditos como de inclusão, e da AM, em que estarão incluídos os erros por omissão, demonstra que, entre as classes preditas, aquela que mais revelou o impacto da substituição dos atributos de terreno pelo novo conjunto de variáveis foi a classe dos Neossolos Litólicos. Esses solos estão localizados na região de encosta do Rebordo do Planalto, locais onde os atributos de terreno utilizados no estudo possuem um grande contraste, principalmente de elevação, declividade e curvaturas.

Possivelmente, com a perda de um terço do poder de explicação das CP em relação às variáveis originais, houve uma perda diferenciada do potencial preditivo dos modelos nas distintas classes de solos preditas. Neste estudo, as variáveis utilizadas estão fortemente relacionadas ao fator de formação relevo, que, por sua vez, é determinante para a distribuição espacial dos Neossolos Litólicos.

As classes de solos pouco representativas na área mapeada não foram espacializadas pelos modelos logísticos. Os solos Plintossolo Argilúvico, Argissolo Acinzentado e Argissolo Vermelho não foram



preditos quando do uso dos atributos de terreno (Figura 2A) e esse fato praticamente não foi alterado com o uso das CP (Figura 2B). Relatos da literatura dão conta de que as RLMM são sensíveis à proporção relativa entre as classes a serem preditas (TEN CATEN et al., 2011). Os modelos logísticos geram os valores de probabilidades baseadas não somente nos valores dos preditores, mas também na proporção relativa entre as classes das variáveis resposta (REAL et al., 2006).

Nas áreas de coxilha, onde predominam os Argissolos, a confusão no mapeamento se deu predominantemente com os Argissolos Vermelho-Amarelos (Figura 2A), quando da utilização dos

atributos de terreno. No entanto, quando foram aplicadas as CP como preditores, o maior percentual de erro nas áreas de terreno relevo suave ondulado a ondulado ocorreu devido à locação nestes locais dos Argissolo Bruno-Acinzentado (Figura 2B). Para os Argissolos Vermelho-Amarelos, a AM caiu de 70,63% para apenas 31,57%, indicando uma variedade de locais onde essa classe está presente no mapa original e deixou de ser predita. Para TEN CATEN et al. (2011), uma melhor predição de classes como dos Argissolos ocorreria a partir de modelos que incorporassem uma variedade maior de covariáveis ligadas ao terreno, ou, ainda, quando informações como a litologia ou uso da terra fossem levados em consideração.

## CONCLUSÃO

O uso de componentes principais como preditores em modelos logísticos pode ser de grande utilidade para a redução do número de variáveis, mas implica perda no potencial preditivo, quando as variáveis originais são naturalmente pouco correlacionadas. As classes de solos foram afetadas de maneira diferenciada pela metodologia, de acordo com a intensidade de sua relação pedogenética com as variáveis predictoras originais. Sendo que classes de solos não representativas na paisagem não são adequadamente espacializadas pelos modelos logísticos múltiplos multinomiais.

## AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de Mestrado concedido ao primeiro autor e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo financiamento e pela Bolsa de Produtividade em Pesquisa concedida ao segundo autor. Agradecem ainda aos revisores científicos pelas contribuições para a melhoria da redação científica deste trabalho.

## REFERÊNCIAS

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

COELHO, F.F.; GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. **Ciência Rural**, v.40, n.10, p.2099-2106, 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-84782010001000008&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782010001000008&lng=en&nrm=iso)>. Acesso em: 20 fev. 2011. doi: 10.1590/S0103-84782010005000156.

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**, v.77, n.1, p.8-18, 2009. Disponível em: <<http://www.sciencedirect.com/science/article/B6VCG-4VCNF40-1/2/6f9e3cc31c08541cb1b162bbeba3cd72>>. Acesso em: 20 fev. 2011. doi:10.1016/j.catena.2008.12.001.

EMBRAPA. **Sistema brasileiro de classificação de solos**. 2.ed. Rio de Janeiro: Embrapa Solos, 2006. 306p.

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://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832008000700023&lng=en&nrm=iso)

06832008000700023&lng=en&nrm=iso>. Acesso em: 20 fev. 2011. doi: 10.1590/S0100-06832008000700023.

GIASSON, E. et al. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. **Scientia Agricola**, v.63, p.262-268, 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-90162006000300008&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162006000300008&lng=en&nrm=iso)>. Acesso em: 20 fev. 2011. doi: 10.1590/S0103-90162006000300008.

HENGL, T.; ROSSITER, D.G. Supervised Landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. **Soil Science Society of America Journal**, v.67, n.6, p.1810-1822, 2003. Disponível em: <<https://www.soils.org/publications/sssaj/abstracts/67/6/1810>>. Acesso em: 20 fev. 2011. doi: 10.2136/sssaj2003.1810

HENGL, T. et al. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. **Geoderma**, v.140, p.417-427, 2007. Disponível em: <<http://www.sciencedirect.com/science/article/B6V67-4NPHMTC-3/2/38c5c93ffa96518abae30ea544c18971>>. Acesso em: 20 fev. 2011. doi:10.1016/j.geoderma.2007.04.022.

KEMPEN, B. et al. Updating the 1:50.000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. **Geoderma**, v.125, p.311-326, 2009. Disponível em: <<http://www.sciencedirect.com/science/article/B6V67-4WD112F-2/2/a3a86e9a30aba03d93856e69f63fa7f5>>. Acesso em: 20 fev. 2011. doi: 10.1016/j.geoderma.2009.04.023.

KIM, K. et al. **Unvaried and multivariate general linear models: theory and applications with SAS**. 2.ed. Boca Raton: Chapman & Hall/CRC, 2007. 549p.

KLAMT, E. et al. **Solos do Município de São Pedro do Sul**. Santa Maria: Departamento de Solos/CCR/UFSM, 2001. 96p.

REAL, R. et al. Obtaining environmental favourability functions from logistic regression. **Environmental and Ecological Statistics**, v.13, n.2, p.237-245, 2006. Disponível em: <<http://dx.doi.org/10.1007/s10651-005-0003-3>>. Acesso em: 20 fev. 2011. doi: 10.1007/s10651-005-0003-3.

SANCHEZ, P.A. et al. Digital soil map of the world. **Science**, v.325, p.680-681, 2009. Disponível em: <<http://www.sciencemag.org/content/325/5941/680>>. Acesso em: 20 fev. 2011. doi: 10.1126/science.1175084.

TEN CATEN, A. et al. Mapeamento Digital: probabilidade associada à distribuição espacial de classes de solos. **Revista Geomática**, v.4, n.2, p.87-97, 2009. Disponível em: <<http://w3.ufsm.br/rgeomatica/page4/08.pdf>>. Acesso em: 20 fev. 2011.

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, 2011. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-06832011000100005&lng=en&nrm=isso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-06832011000100005&lng=en&nrm=isso)>. Acesso em 16 maio, 2011. doi: 10.1590/S0100-06832011000100005.