

Estatística multivariada aplicada à diminuição do número de preditores no mapeamento digital do solo

Alexandre ten Caten⁽¹⁾, Ricardo Simão Diniz Dalmolin⁽²⁾, Fabrício Araújo Pedron⁽²⁾
e Maria de Lourdes Mendonça-Santos⁽³⁾

⁽¹⁾Instituto Federal Farroupilha, Campus Júlio de Castilhos, São João do Barro Preto, s/nº, CEP 98130-000 Júlio de Castilhos, RS. E-mail: acaten@yahoo.com.br ⁽²⁾Universidade Federal de Santa Maria, Departamento de Solos, Avenida Roraima, nº 1.000, CEP 97105-900 Santa Maria, RS. E-mail: dalmolin@cnpq.br, fapedron@gmail.com ⁽³⁾Embrapa Solos, Rua Jardim Botânico, nº 1.024, CEP 22460-000 Rio de Janeiro, RJ. E-mail: loumendonca@cnpq.embrapa.br

Resumo – O objetivo deste trabalho foi avaliar a possibilidade de se gerar um menor conjunto de preditores não correlacionados e potencialmente aplicáveis ao mapeamento digital de solos, pelo uso da estatística multivariada. Os atributos de terreno, elevação, declividade, distância à drenagem, curvatura planar, curvatura de perfil, radiação relativa disponível, logaritmo natural da área de contribuição, índice de umidade topográfica e capacidade de transporte de sedimento, foram transformados pelo método Varimax nas variáveis: altimetria, hidrologia e curvatura. Essa transformação representou uma concentração de 65,57% da variabilidade dos dados originais nas três novas componentes. As novas variáveis possibilitam o emprego de menor quantidade de dados nos modelos preditivos, além do fato de serem não correlacionados. A rotação Varimax permite que a relação com o ambiente de formação do solo seja explicitamente inserida nos modelos preditivos.

Termos para indexação: análise de componentes principais, atributos de terreno, pedometria, sensoriamento remoto.

Multivariate analysis applied to reduce the number of predictors in digital soil mapping

Abstract – The objective of this work was to assess the possibility of generating a smaller set of uncorrelated predictors, potentially applicable to digital soil mapping, by multivariate statistical analysis. The terrain attributes, elevation, slope, stream distance, planar curvature, profile curvature, relative available radiation, natural logarithm of the contributing area, topographic wetness index, and sediment transport capacity, were transformed by the Varimax method into the variables: altimetry, hydrology, and curvature. This transformation represented a variability concentration of 65.57% of the original data in the three new components. The new variables enable the use of a smaller amount of data set in prediction models, besides the fact that they are uncorrelated. Varimax rotation allows the relationship between environment and soil formation to be explicitly included in the prediction models.

Index terms: principal components analysis, terrain attributes, pedometry, remote sensing.

Introdução

Diversos avanços tecnológicos têm possibilitado desenvolver uma variedade de técnicas de observação do ambiente e estimulado estudos ligados à distribuição espacial de classes e propriedades de solo. Os avanços em sensoriamento remoto (SR) têm ocorrido rapidamente, principalmente se for considerado o aumento das resoluções espacial e espectral, ambas de grande interesse para o mapeamento digital de solos (MDS). Com isso, fatores e processos de formação do solo podem ser modelados a partir de variáveis mensuradas por técnicas de SR, o que possibilita a obtenção de dados de todas as variáveis do modelo Scorpan (McBratney et al., 2003).

Embora o MDS seja beneficiado pela grande gama de preditores disponíveis, ainda são encontradas dificuldades na sua realização, uma vez que é necessário determinar as variáveis de maior e melhor potencial preditivo, assim como conhecer suas interrelações e efeitos de multicolinearidade. Ao se considerar um problema comum em regressões, a multicolinearidade manifesta-se quando as variáveis independentes possuem fortes relações lineares. A ocorrência de multicolinearidade leva a elevados erros-padrão na predição dos parâmetros do modelo ou, até mesmo, inviabiliza o seu ajuste (Manly, 2008). A ausência de multicolinearidade seria uma das premissas para o estabelecimento de um modelo adequado (Hengl et al., 2007).

Entre as metodologias de predição espacial de propriedades e classes de solo está a regressão logística múltipla (RLM) (Giasson et al., 2006; Hengl et al., 2007; Kempen et al., 2009), que estima a probabilidade de ocorrência da variável dependente a partir de variáveis independentes, as quais podem ser discretas, contínuas ou binárias. No entanto, um dos problemas de aplicação dessa técnica de modelos lineares generalizados é que não são verificados efeitos de multicolinearidade entre os dados na fase de ajuste dos modelos (Hengl et al., 2007), sendo essa a maior limitação desses modelos, quando aplicados à predição de propriedades e classes de solos (Kempen et al., 2009).

O emprego de transformações nos dados originais pode torná-los não correlacionados e evitar o efeito da multicolinearidade. Neste caso, a análise de componentes principais (ACP) é uma possibilidade, uma vez que reduz o número de variáveis iniciais – o que a torna mais facilmente manipulável – e gera novas variáveis não correlacionadas. Conforme Webster (2001), a ACP constitui-se em uma transformação de dados, com rotação rígida dos dados originais para um novo conjunto de eixos por meio de procedimento matemático, não estatístico. Essa análise não requer pressuposição da normalidade da distribuição, não realiza testes estatísticos de significância, e pode ser usada para reduzir o número de variáveis (Vitharana et al., 2008), identificar padrões (Gomes et al., 2004) e gerar variáveis não correlacionadas (Hengl et al., 2007). As componentes principais (CP), quando utilizadas como variáveis independentes em modelos preditivos de propriedades e classes de solos, são definidas como componentes predictoras do solo (Hengl et al., 2004).

Na década de 1970, Webster & Burrough (1972) utilizaram a ACP para selecionar, entre um conjunto de variáveis, as que explicavam a maior parte da variabilidade dos dados, para analisar a coerência e a homogeneidade dos delineamentos de classe de solos. A aplicação de ACP pode apresentar desvantagens, como a transformação dos dados originais em novas variáveis sem significado físico, químico ou biológico, e o fato de ter interpretação mais complexa. É importante estar atento e associar conhecimentos próprios às novas informações geradas, para lhes conferir significado pedológico e evitar a correlação (redundância) entre os preditores, o que irá garantir a estabilidade dos modelos (Webster, 2001).

O objetivo deste trabalho foi avaliar a possibilidade de se gerar um menor conjunto de preditores não correlacionados aplicáveis ao mapeamento digital de solos, pelo uso da estatística multivariada.

Material e Métodos

A área de estudo localiza-se no Município de São Pedro do Sul, Região Central do Estado do Rio Grande do Sul, entre as coordenadas 29°46'–29°26'S e 54°30'–53°56'W. A área abrange uma região transitória entre o Planalto Médio e a Depressão Central do Rio Grande do Sul (RS), de onde se obtiveram dados de um levantamento convencional de solos, por meio de foteointerpretação e verificação a campo das classes de solos presentes. A região mapeada pela metodologia convencional é denominada área de referência e encontra-se detalhadamente descrita em ten Caten et al. (2011a).

Para a derivação dos atributos do terreno a serem usados no estudo, foi utilizado o modelo digital de elevação (MDE), obtido a partir de dados do "Shuttle Radar Topography Mission" (SRTM) (Rabus et al., 2003). Os atributos de terreno elevação (ELEV), declividade (DECL), distância à drenagem (DIST), curvatura planar (CPLN), curvatura de perfil (CPRF), radiação relativa disponível (RADI), logaritmo natural da área de contribuição (LNAC), índice de unidade topográfica (IUT) e capacidade de transporte de sedimento (CTS) foram gerados de acordo com procedimento detalhado por ten Caten et al. (2011b). Um conjunto de 70.000 pontos (20% dos pixels, com 50 m de resolução espacial) foram aleatoriamente gerados para amostrar os planos de informação (PI) dos nove atributos de terreno e do mapa de solos, para possibilitar a tabulação dos dados em formato texto (ASCII) e a análise estatística. As 70.000 amostras foram distribuídas proporcionalmente na área de cada classe de solo.

A adequabilidade das amostras para aplicação da ACP foi verificada pelo teste individual "measure of sample adequacy" (MAS) e geral "Kaiser Meyer Olkin" (KMO), de acordo com Gobin et al. (2000). A ACP foi realizada a partir de dados padronizados (média igual a zero e variância igual a um). Como a seleção das variáveis que irão fazer parte dos modelos logísticos deve ser feita com critérios pedológicos – e a ACP é uma transformação ortogonal dos dados originais –, o estabelecimento de relações pedogenéticas entre CP

e o solo fica comprometido (Hengl et al., 2004). Para contornar essa limitação, foi aplicada a rotação Varimax (Hair et al., 2007) nos CP, a qual potencializa as relações entre as variáveis originais e as novas variáveis. O método de rotação Varimax foi aplicado aos autovetores para maximizar a soma da variância do quadrado dos pesos, pois um novo conjunto de pesos com valores grandes (próximo a 1) ou pequenos (próximo a 0) facilita a interpretação dos CP. Para a realização dos cálculos, foi utilizado o SAS (Kim, 2007).

Os nove atributos de terreno e as CP, geradas pelo método Varimax, foram utilizados como variáveis preditoras em RLM. As classes de solos, definidas pelo levantamento convencional na área de referência, foram utilizadas como variáveis resposta nos modelos logísticos. Foram utilizadas as classes de solos no nível de subordem (segundo nível categórico) do Sistema Brasileiro de Classificação de Solos (Santos et al., 2006). Em todos os modelos logísticos testados, foram considerados apenas os coeficientes significativos, a 5% de probabilidade, pelo teste de Wald (Chatterjee & Hadi, 2006). Os modelos gerados possibilitaram a produção de mapas de solos com e sem a aplicação das CP, os quais foram comparados pelo índice kappa (Congalton, 1991). Todas as etapas de geoprocessamento dos dados foram executadas no programa ArcGIS 9.3 (Environmental Systems Research Institute, 2008).

Resultados e Discussão

Nas imagens da distribuição espacial dos atributos do terreno derivados do MDE foram verificadas duas regiões bem distintas, a da Encosta do Planalto e a da Depressão Central, que exercem influência decisiva na gênese dos solos da área (Figura 1). Na Encosta do Planalto, de relevo forte ondulado a montanhoso, predominam os processos de remoção constante do solo, que condiciona a formação de solos mais rasos. Na Depressão Central, de relevo plano a suave ondulado, ocorre o processo de acumulação e permanência constante de água.

Cada um dos atributos de terreno apresenta influência específica na gênese dos solos da área. Na parte superior da Figura 1 A, observa-se a predominância de maiores elevações do terreno. Em altitudes de até 446,25 m predominam os Neossolos e Cambissolos, uma vez que nesses locais o relevo montanhoso, com declividade superior a 30° (57%) (Figura 1 B), possibilita a remoção constante do solo. Em posições de relevo ondulado com

declividades próximas a 10° (17%), na região central do Município de São Pedro do Sul, ocorre a infiltração de água mais profunda no solo, e possibilita a formação dos Argissolos. Áreas onde a presença constante de água sobre a paisagem (tons claros da Figura 1 H) potencializa os processos de oxirredução e gleização e proporcionam a ocorrência de Planossolos e Gleissolos.

Os dados de atributos do terreno não apresentam distribuição normal, o que é preconizado para maioria das análises estatísticas que utilizam parâmetros de tendência central. Isso pode ser visualizado no baixo contraste das imagens na Figura 1 (predomínio de tons muito escuros ou muito claros), e na forma assimétrica de alguns dos histogramas dessa figura. Contudo, a pressuposição da normalidade da distribuição não é necessária para a aplicação de ACP (Gobin et al., 2000).

Para a aplicação da ACP na matriz de correlação, foi verificada a adequabilidade dos dados pelos testes individual de MAS e geral KMO (Gobin et al., 2000). Entre os atributos do terreno CPRF e LNAC foi obtido valores MAS de 0,58 e 0,56, respectivamente, que podem ser considerados baixos. O valor de KMO do conjunto de todos os atributos foi de 0,66, e reflete predomínio de baixas correlações entre os atributos de terreno; contudo, esse valor não inviabiliza a aplicação da ACP ao conjunto de dados. Vitharana et al. (2008) utilizaram CP para identificar atributos do terreno que se relacionassem a zonas de manejo em agricultura de precisão, e atribuíram valor KMO de 0,67, muito similar ao encontrado no presente trabalho. Apesar disso, os autores relatam que a metodologia com uso de CP auxilia no adequado delineamento das zonas de manejo, a partir de atributos do terreno.

Após a aplicação da ACP, aos nove atributos do terreno, foram geradas nove CP, cada uma com um percentual decrescente da variabilidade dos dados iniciais (Figura 2). Os autovalores foram superiores a 1, para as três primeiras CP (Figura 2 A). Como a ACP foi realizada a partir de dados padronizados – média igual a zero e variância igual a 1 –, apenas são significativas as CP com valores de autovalor superior a unidade, uma vez que esse já é o valor da variância de cada uma das variáveis individualmente. No presente trabalho, esse resultado justificaria o uso da ACP para a redução do número de variáveis e aumento do poder de discriminação de cada uma das novas variáveis (CP).

Quanto ao percentual de variância explicado pelas novas variáveis, verifica-se que a primeira componente é

responsável por um terço da variância dos dados originais (33,00%). A segunda e a terceira componentes juntas têm um poder de explicação equivalente ao da primeira componente, em torno de um terço da variabilidade total (32,57%). Dessa forma, as três primeiras componentes já explicam, sozinhas, 65,57% da variância dos dados. Portanto, a sua utilização possibilitaria a simplificação do número de variáveis, embora com a perda de aproximadamente 34% do padrão de variabilidade dos dados. As novas variáveis (CP) não possuem nenhum significado pedológico, pois são produtos da

transformação ortogonal dos dados originais. Portanto, as relações diretas feitas entre os atributos do terreno (Figura 1) e os tipos de solos da região, por meio do comportamento da paisagem frente ao transporte de sedimentos, exposição à radiação ou direcionamento dos fluxos de água na paisagem, não são aplicáveis.

Hengl & Rossiter (2003) utilizaram ACP para mitigar o efeito de multicolinearidade entre atributos do terreno. E identificaram que as cinco primeiras componentes foram responsáveis por cerca de 80% da variabilidade, de um conjunto de nove atributos do terreno. Esses autores

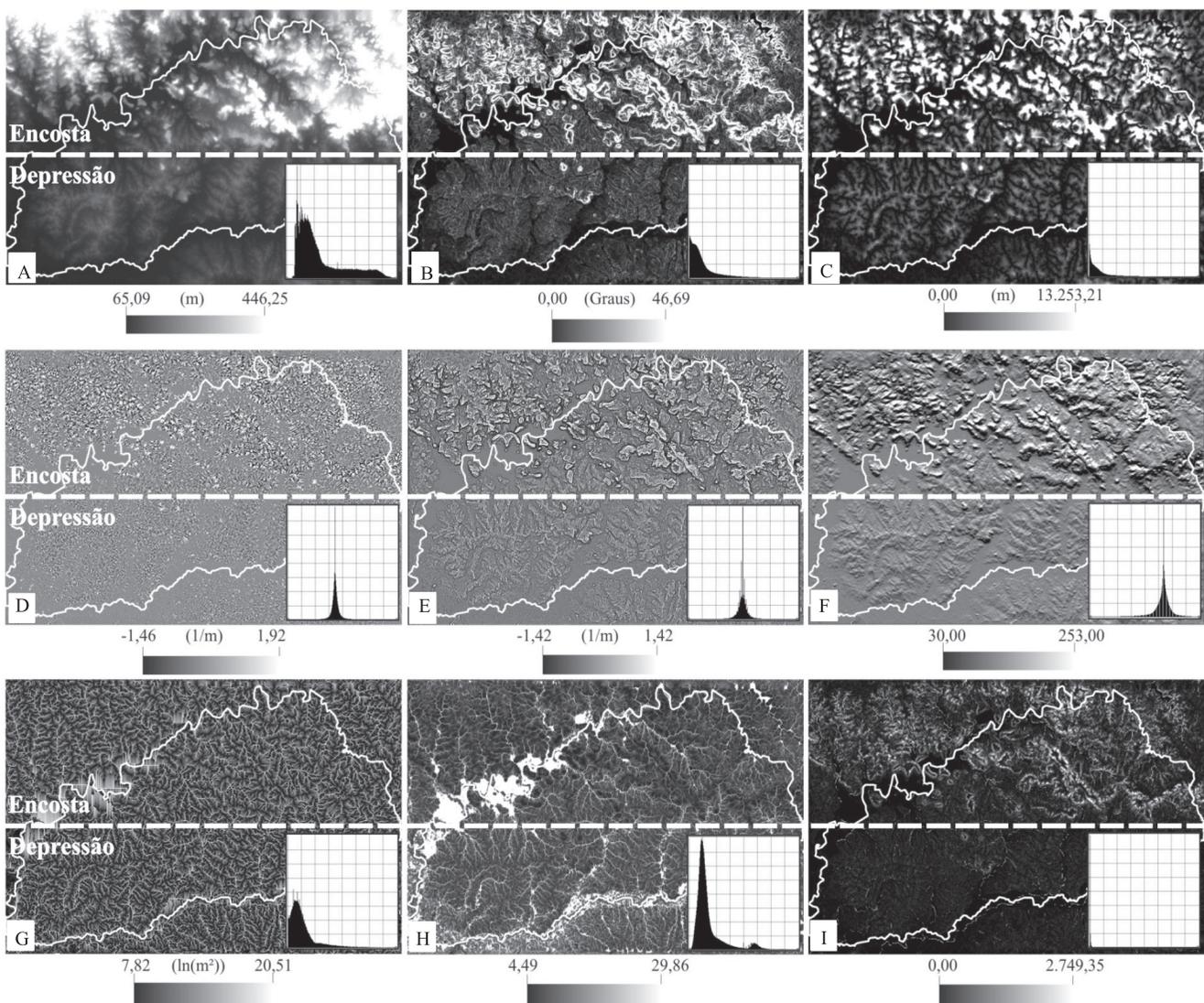


Figura 1. Distribuição espacial dos atributos do terreno (e seus histogramas), na área de estudo: A, elevação (ELEV); B, declividade (DECL); C, distância à drenagem (DIST); D, curvatura planar (CPLN); E, curvatura de perfil (CPRF); F, radiação relativa disponível (RADI); G, área de contribuição (LNAC); H, índice de umidade topográfica (IUT); I, capacidade de transporte de sedimento (CTS). Nas figuras, a área superior está localizada na Encosta do Planalto e a parte inferior na Depressão Central.

também aplicaram a ACP aos atributos do terreno, com divisão dos dados em duas sub-regiões: relevo de planície e de montanha. As proporções da variância explicada pelas CP praticamente não se alteraram, sendo que, até a terceira componente, em ambas as áreas, por volta de 70% da variabilidade era retida. Esses resultados são semelhantes ao da Figura 2 A, em que as cinco primeiras componentes acumulam cerca de 84,14% da variância.

A Figura 2 B, com imagem do “scree plot”, também pode ser usada para a verificação da importância das componentes mais elevadas na explicação da variância. Esse gráfico pode ser utilizado para a tomada de decisão de quantas componentes devem ser retidas após a ACP. Para uma aplicação mais efetiva da ACP, esperava-se que a reta tracejada, que intercepta essas componentes, fosse mais próxima e paralela ao eixo

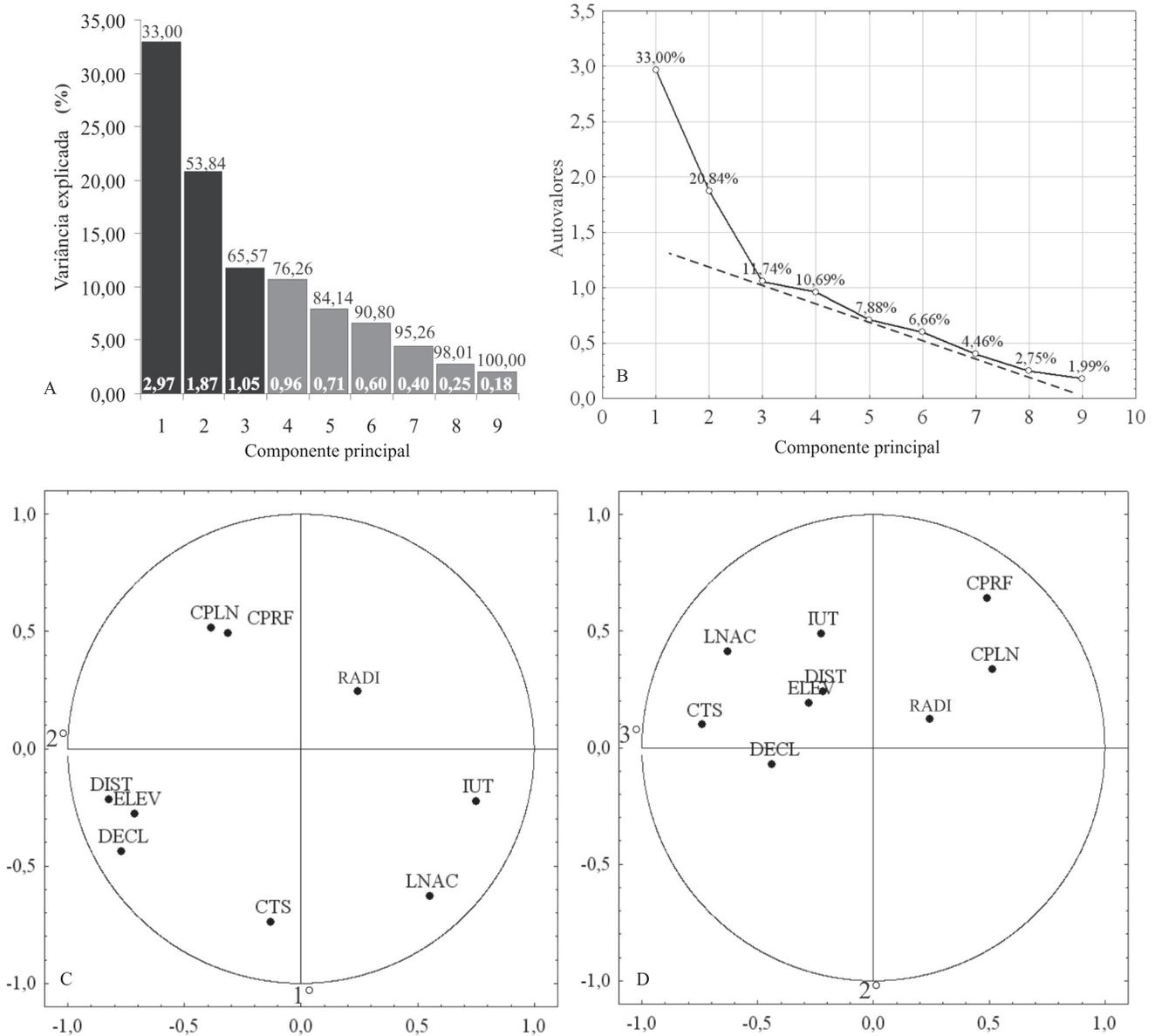


Figura 2. Variância explicada pelas nove componentes principais (CP) geradas (A), "Scree plot" (B), diagramas de dispersão circular (C e D). ELEV, elevação; DECL, declividade; CPLN, curvatura planar; CPRF, curvatura de perfil; DIST, distância à drenagem; RADI, radiação relativa disponível; LNAC, área de contribuição; IUT, índice de umidade topográfica; e CTS, capacidade de transporte de sedimento. Autovalores de cada componente constam em branco na base e no interior de cada barra. As três primeiras barras em negrito possuem autovalor maior do que a unidade. Valor sobre as barras indica o percentual de variância acumulada.

horizontal. O método "scree plot" pode ser substituído pelo método dos autovalores com valor superior a unidade, ou ainda, pelo conjunto de CP que expliquem acima de 70% da variância, como critérios para decisão do número de CP a serem retidas (Manly, 2008).

Os pesos dos nove atributos do terreno, nas três primeiras componentes retidas, podem ser visualizados nos diagramas de ordenação unitários das Figuras 2 C e D. Os atributos mais significativos, para os 33% da variabilidade explicada pelo primeiro eixo, são DIST, DECL e ELEV, que são correlacionados entre si, como indica a sua proximidade dentro do círculo unitário. Da mesma forma, a IUT tem correlação positiva com a primeira CP. Devido à posição praticamente simétrica dessas quatro variáveis no diagrama de ordenação unitária, pode-se dizer que elas são inversamente correlacionadas. Os demais atributos do terreno possuem pequena contribuição na variabilidade dos dados dessa primeira componente.

As contribuições dos atributos do terreno na segunda componente, com variância acumulada de 22,84%, podem ser visualizadas nas projeções desses atributos para a linha vertical do círculo unitário (Figura 2 C). Os atributos mais significativos são a CTS e a LNAC, embora ainda com correlações bastante baixas com essa componente. Os valores de CPLN e CPRF mostraram essas variáveis bastante correlacionadas entre si, nessas duas primeiras componentes. Na Figura 2 D estão as correlações dos atributos do terreno com a segunda e terceira componentes. Todos os atributos encontram-se menos dispersos e mais próximos ao centro do círculo unitário, o que era de se esperar, visto que o poder de explicação das variáveis diminui à medida que é menor a variância ainda retida pelas componentes mais elevadas. Em todas as três primeiras componentes, o atributo RADI encontra-se próximo ao centro do círculo unitário, o que é indicativo de sua pouca relação com as novas variáveis (CP) geradas.

Após a aplicação da rotação Varimax, obteve-se as três componentes que constituem os componentes preditores do solo deste estudo (Tabela 1). A primeira componente foi denominada componente altimetria, por ter maior relação com ELEV, DECL e DIST. A segunda, foi denominada componente hidrológica, devido aos valores mais significativos de correlações com as variáveis ligadas ao movimento da água na paisagem, como LNAC e IUT. A terceira foi denominada componente curvatura, devido aos

maiores pesos dos atributos CPLN e CPRF. A rotação dos pesos potencializou a relação entre os atributos de terreno originais e cada uma das componentes. Assim, a falta de significado pedológico das CP inicialmente geradas fica parcialmente contornada, pois a cada uma pode-se associar as variáveis originais (atributos do terreno). Abordagem similar, na busca por um sentido pedológico para as CP utilizadas, foi adotada por Park & Burt (2002). Em estudo que relacionou processos de formação e atributos de terreno, os autores associaram cinco CP a fatores pedogeomorfológicos, os quais foram a podzolização, dinâmica de nutrientes, lixiviação de Mn, intemperização e acidificação do solo.

A proporção da variância, de cada atributo, explicada pelas três primeiras componentes, pode ser visualizada na coluna das comunalidades (Tabela 1). Valores próximos a um indicam que o atributo foi bem contemplado pelas componentes retidas. Valores baixos, como 0,52 (CPLN) e até mesmo 0,13 (RADI), indicam que estas variáveis foram menos contempladas pelo novo conjunto de variáveis. Os baixos valores de comunalidades estão relacionados com o fato de que as três primeiras componentes juntas explicam apenas 65,57% da variabilidade dos dados originais (Figura 2 A). Sendo que mais de um terço da variância dos atributos originais será perdida com a nova estrutura de dados.

Valores não significativos dos interceptos e coeficientes não foram utilizados para gerar os modelos logísticos (Tabela 2). Segundo Kempen et al. (2009),

Tabela 1. Pesos dos nove atributos do terreno nas três primeiras componentes principais (CP) retidas após a rotação por Varimax.

Atributo de terreno	Componente principal			Comunalidade
	1	2	3	
ELEV	<u>0,756</u>	-0,074	0,218	0,63
DECL	<u>0,870</u>	-0,184	-0,053	0,79
DIST	<u>0,810</u>	-0,131	0,327	0,78
CPLN	0,038	-0,264	<u>0,673</u>	0,52
CPRF	0,027	-0,019	<u>0,865</u>	0,75
RADI	-0,325	0,062	0,151	0,13
LNAC	-0,029	<u>0,904</u>	-0,231	0,87
IUT	-0,419	<u>0,822</u>	-0,010	0,85
CTS	0,549	0,420	-0,309	0,57

⁽¹⁾A comunalidade indica o total de variabilidade de cada variável explicada pelas três componentes retidas. ELEV, elevação; DECL, declividade; CPLN, curvatura planar; CPRF, curvatura de perfil; DIST, distância à drenagem; RADI, radiação relativa disponível; LNAC, área de contribuição; IUT, índice de umidade topográfica; e CTS, capacidade de transporte de sedimento. Valores sublinhados apresentam peso superior a 0,600.

Tabela 2. Valores dos interceptos e coeficientes dos modelos logísticos para o segundo nível categórico gerados a partir das covariáveis atributos de terreno e componentes principais.

Variável	CX ⁽¹⁾				
	SH	PBAC	PV	RL	PVA
Atributos de terreno					
Intercepto	36,384	26,510	17,125	-0,278 ^{ns}	20,460
ELEV	-0,077	-0,053	-0,027	0,014	-0,045
DECL	-0,238	-0,031	-0,027	-0,050 ^{ns}	-0,198
DIST	-0,001	-0,001	0,000	0,001	0,000
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 ^{ns}	-0,026
LNAC	-1,572	-1,472	-1,383	-0,311 ^{ns}	-1,421
IUT	1,096	1,077	0,989	0,084 ^{ns}	0,835
CTS	0,007 ^{ns}	-0,006 ^{ns}	-0,009 ^{ns}	0,010 ^{ns}	0,016 ^{ns}
Componentes principais					
Intercepto	-1,948	1,967	2,963	-1,660	-0,118 ^{ns}
Altimetria	-0,096	-0,062	-0,041	0,003 ^{ns}	-0,074
Hidrologia	-0,068	-0,074	-0,068	0,040	-0,111
Curvatura	0,202	0,119	0,073	0,021	0,139

⁽¹⁾Classe de referência. CX, Cambissolo Háplico; PV, Argissolo Vermelho; PVA, Argissolo Vermelho-Amarelo; RL, Neossolo Litólico; PBAC, Argissolo Bruno-Acinzentado; e SH, Solo Hidromórfico. ELEV, elevação; DECL, declividade; CPLN, curvatura planar; CPRF, curvatura de perfil; DIST, distância à drenagem; RAD, radiação relativa disponível; LNAC, área de contribuição; IUT, índice de umidade topográfica; e CTS, capacidade de transporte de sedimento.

^{ns}Não significativo, a 5% pelo teste de Wald.

preditores não significativos não são correlacionados com as classes de solo a serem preditas e, portanto, devem ser excluídos da fase de implementação dos modelos. A partir desses parâmetros, para cada classe de solo, exceto a classe de referência, obtém-se um modelo logístico, o qual possibilita prever a probabilidade de determinada classe de solo ser encontrada em um dado local da paisagem (ten Caten et al., 2011b).

A classe dos Solos Hidromórficos (Associação Gleissolos Melânicos + Neossolos Quartzarênicos) não foi predita por nenhum dos modelos empregados (Figura 3 A e B). Nos locais onde essa classe havia sido mapeada, foram preditos, predominantemente, Argissolos Bruno-Acinzentados. Segundo ten Caten et al. (2011b), o potencial preditivo de modelos logísticos pode ser melhorado se maior variedade de fatores de formação do solo forem utilizados.

A aplicação dos atributos de terreno como preditores resultou em um índice kappa de 46,12%, quando o mapa digital foi comparado ao mapa convencional da área de referência, utilizado para treinar os modelos. Esse valor é dez pontos percentuais superior ao reportado por Giasson et al. (2006), e levemente

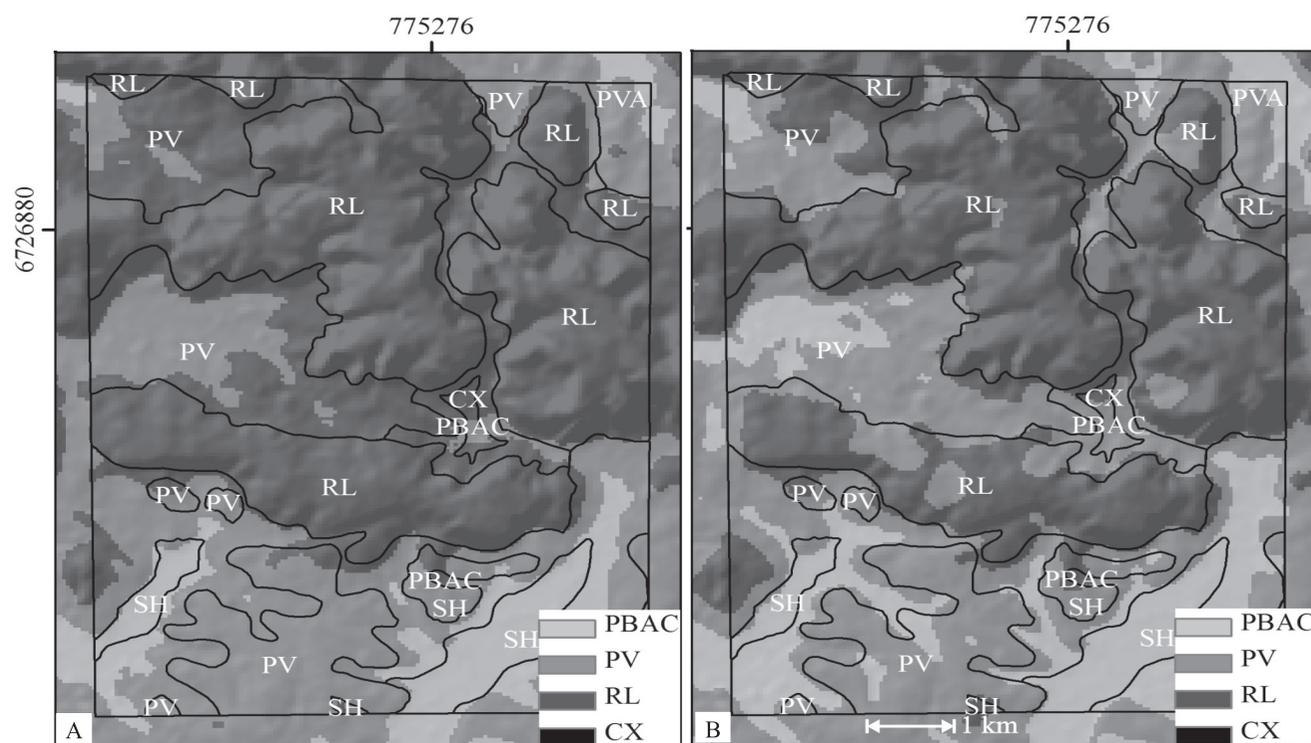


Figura 3. Mapa de solos utilizando atributos de terreno como preditores (A), e mapa de solos utilizando como preditores os componentes preditores do solo (B). Sobreposto às imagens está o mapa de solos convencional. CX, Cambissolo Háplico; PV, Argissolo Vermelho; PVA, Argissolo Vermelho-Amarelo; RL, Neossolo Litólico; PBAC, Argissolo Bruno-Acinzentado; e SH, Solo Hidromórfico.

superior ao índice kappa encontrado por Hengl & Rossiter (2003), de 42,6 %. Contudo, esses baixos valores de reprodutibilidade dos mapas convencionais pela abordagem digital reforçam a necessidade de que mais pesquisas sejam realizadas, para identificar as melhores metodologias e os preditores mais adequados para capturar a complexidade inerente à distribuição espacial das classes de solos.

O emprego das componentes altimetria, hidrológica e curvatura possibilitou gerar um mapa de solos com reprodutibilidade de 43,6% (índice kappa), em relação ao mapa convencional. Esse valor é muito próximo ao encontrado quando nove atributos de terreno foram usados como preditores nos modelos (46,12%). Com isso, mesmo que a aplicação dos componentes preditores do solo implique em uma variabilidade cerca de um terço menor do que a atual, com a aplicação dos nove atributos de terreno, as RLM simplificadas mostraram potencial preditivo similar aos modelos completos.

A semelhança observada entre os índices kappa pode ser explicada, em parte, pela predominância de solos com a pedogênese ditada pelo relevo, na área de estudo, e a utilização unicamente de preditores ligados ao fator de formação relevo. A influência de uma transformação ortogonal (ACP), em um conjunto de preditores ligados a distintos fatores de formação do solo, como organismos, material de origem, clima e tempo, e de como essa transformação irá afetar o potencial preditivo das novas variáveis geradas (CP) ainda deve ser investigada.

Conclusões

1. Para a área estudada, há três componentes preditores do solo: altimetria, hidrologia e curvatura, que representam a forte influência do fator relevo como condicionante dos fluxos de água que atuam na gênese do solo.

2. O emprego de componentes preditores do solo possibilita a geração de modelos simplificados e com potencial preditivo similar aos que empregam todo o conjunto de variáveis originais, o que permite gerar mapas de solos com índices de acerto similares aos reportados na literatura.

3. A rotação Varimax possibilita que a relação com o ambiente de formação do solo seja explicitamente inserida nos modelos preditivos a serem propostos.

Referências

- CATEN, A. ten; DALMOLIN, R.S.D.; PEDRON, F. de A.; MENDONÇA-SANTOS, M. de L. Extrapolação das relações solo-paisagem a partir de uma área de referência. *Ciência Rural*, v.41, p.812-816, 2011a.
- CATEN, A. ten; DALMOLIN, R.S.D.; PEDRON, F. de A.; MENDONÇA-SANTOS, M. de L. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. *Revista Brasileira de Ciência do Solo*, v.35, p.53-62, 2011b.
- CHATTERJEE, S.; HADI, A.S. **Regression analysis by example**. 4th ed. New York: J. Willey, 2006. 375p.
- CONGALTON, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v.37, p.35-46, 1991.
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE. **ArcGIS**. Version 9.3. Redlands: ESRI, 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. *Scientia Agricola*, v.63, p.262-268, 2006.
- GOBIN, A.; CAMPLING, P.; DECKERS, J.; FEYEN, J. Quantifying soil morphology in tropical environments: methods and application in soil classification. *Soil Science Society of America Journal*, v.64, p.1423-1433, 2000.
- GOMES, J.B.V.; CURTI, N.; MOTTA, P.E.F.; KER, J.C.; MARQUES, J.J.G.S.M.; SCHULZE, D.G. Análise de componentes principais de atributos físicos, químicos e mineralógicos de solos do bioma cerrado. *Revista Brasileira de Ciência do Solo*, v.28, p.137-153, 2004.
- HAIR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise multivariada de dados**. 5.ed. Porto Alegre: Bookman, 2007. 593p.
- HENGL, T.; HEUVELINK, G.B.M.; STEIN, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, v.120, p.75-93, 2004.
- 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, p.1810-1822, 2003.
- HENGL, T.; TOOMANIAN, N.; REUTER, H.I.; MALAKOUTI, M.J. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma*, v.140, p.417-427, 2007.
- 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*, v.151, p.311-326, 2009.
- KIM, K.; TIMM, N.; BALAKRISHNAN, N. **Unvaried and multivariate general linear models**: theory and applications with SAS. 2nd ed. New York: Chapman and Hall: CRC, 2007. 549p.

- MANLY, B.J.F. **Métodos estatísticos multivariados**: uma introdução. 3.ed. Porto Alegre: Bookman, 2008. 229p.
- MCBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. **Geoderma**, v.117, p.3-52, 2003.
- PARK, S.J.; BURT, T.P. Identification and characterization of pedogeomorphological processes on a hillslope. **Soil Science Society of America Journal**, v.66, p.1897-1910, 2002.
- RABUS, B.; EINEDER, M.; ROTH, A.; BAMLER, R. The shuttle radar topography mission - a new class of digital elevation models acquired by spaceborne radar. **Journal of Photogrammetry and Remote Sensing**, v.53, p.241-262, 2003.
- SANTOS, H.G. dos; JACOMINE, P.K.T.; ANJOS, L.H.C. dos; OLIVEIRA, V.A. de; OLIVEIRA, J.B. de; COELHO, M.R.; LUMBRERAS, J.F.; CUNHA, T.J.F. (Ed.). **Sistema brasileiro de classificação de solos**. 2.ed. Rio de Janeiro: Embrapa Solos, 2006. 306p.
- VITHARANA, U.W.A.; MEIRVENNE, M. van; SIMPSON, D.; COCKX, L.; BAERDEMAEKER, J.D. Key soil and topographic properties to delineate potential management classes for precision agriculture in the European loess area. **Geoderma**, v.143, p.206-215, 2008.
- WEBSTER, R. Statistics to support soil research and their presentation. **European Journal of Soil Science**, v.52, p.331-340, 2001.
- WEBSTER, R.; BURROUGH, P.A. Computer-based soil mapping of small areas from sample data. 1. Multivariate classification and ordination. **Journal of Soil Science**, v.23, p.210-221, 1972.

Recebido em 16 de novembro de 2010 e aprovado em 7 de abril de 2011