

Uso de componentes principais na determinação de variáveis preditoras aplicadas ao mapeamento digital de solos

Ândrea Machado Pereira Franco⁽²⁾; Ricardo Simão Diniz Dalmolin⁽³⁾; Luis Fernando Chimelo Ruiz⁽⁴⁾; Priscila Vogelei Ramos⁽⁵⁾

⁽¹⁾ Trabalho executado com recursos da CAPES

⁽²⁾ Pós-Doutoranda, Bolsista PNPd/CAPES, Departamento de Solos (DS), Universidade Federal de Santa Maria (UFSM); Santa Maria – RS, email: pfranco20@hotmail.com; ⁽³⁾ Professor Associado DS/UFSM; ⁽⁴⁾ Tecnólogo em Geoprocessamento, bolsista DTI-C/CNPq; Universidade Federal de Santa Catarina, Campus Curitibanos, SC; ⁽⁵⁾ Mestranda, Bolsista CNPq, DS/UFSM.

RESUMO: O Mapeamento Digital de Solos (MDS) é uma técnica capaz de criar e disponibilizar informações sobre o solo, com métodos tradicionais associados a sistemas de inferência baseados nos fatores de formação do solo. O objetivo deste estudo foi conhecer as variáveis que mais influenciam na predição de classes de solos através da análise de componentes principais (ACP), reduzindo o número de variáveis que são aplicadas ao MDS. O estudo foi realizado na sub-bacia do Rio Vacacaí-Mirim na região do Rebordo do Planalto do RS. A variável ELEV (elevação) foi originada da interpolação das curvas de nível da carta topográfica. Posteriormente foram obtidas as variáveis ambientais também chamadas de atributos do terreno: CPLAN, CPRF, LS, WET, DECL, DCOT, HIDRO, ILUS e DPNDVI. Os componentes principais (CP's) e as correlações foram obtidas utilizando o software livre R. Entre os dois grupos de variáveis que se correlacionam e influenciaram nos maiores valores de variância, pode-se excluir as variáveis LS, DPNDVI e HIDRO, pois essas que não são consideradas boas preditoras de classes de solos. A variáveis ILUS também não seria significativa para a predição de classes de solos em função da que a mesma somente tem melhor influência na variabilidade da PC 4. Obtivemos uma boa predição com CPLAN, CPRF, DECL, ELEV, WET e DCOT. Com a ACP é possível reduzir e prever o conjunto de variáveis ambientais que não se correlacionam entre si e selecionar variáveis preditoras de classes de solos aplicadas ao MDS.

Termos de indexação: atributos do terreno, MDS, variáveis ambientais.

INTRODUÇÃO

A distribuição espacial das classes de solos e seus respectivos atributos de formação são de grande valia, pois contribuem para minimizar possíveis problemas ambientais decorrentes principalmente de ações antrópicas. A utilização de Mapeamento Digital dos Solos (MDS) colabora com o desenvolvimento desses estudos (McBratney et al., 2003).

O MDS está intimamente relacionado com a geração de um sistema de informações que estabelecem relações matemáticas entre variáveis ambientais e as classes de solos, podendo consequentemente prever a distribuição espacial dessas classes (Coelho & Giasson, 2010).

Para tanto é indispensável que seja criado um banco de dados com variáveis ambientais que mais influenciam na variabilidade de solos do local de interesse, para posteriormente desenvolver novos trabalhos na área do MDS.

Uma melhor representação das relações solo-paisagem e das condições ambientais de formação do solo é feita utilizando alguns atributos do terreno e índices espectrais, também chamados de variáveis ambientais. Trabalhos de ten Caten et al. (2009; 2011a; 2011b) e Coelho & Giasson (2010) mostram que os atributos do terreno originários do mapeamento digital de elevação (MDE) mais utilizados são: declividade, curvatura planar, curvatura de perfil, distância à drenagem, umidade topográfica, fator topográfico e radiação relativa disponível. Esses atributos são melhores detalhados em ten Caten et al. (2009).

A incidência da radiação solar sobre as vertentes, juntamente com o regime de distribuição das chuvas de uma determinada região irão influenciar sobre as respostas dos índices espectrais dos alvos, principalmente àqueles que avaliam a vegetação (Perez et al., 2004).

Estas variáveis ambientais, consequentemente são influenciadas pelo aumento da temperatura do solo e diminuição do teor de umidade, pois essas variáveis podem influenciar na evapotranspiração potencial, podendo causar algumas alterações nas propriedades químicas e físicas do solo (Corrêa, 2008).

Outra variável que vem sendo frequentemente avaliada é o NDVI (*Normalized Difference Vegetation Index*), proposto por Rouse et al. (1973), utilizado na caracterização e no monitoramento da vegetação, bem como na detecção e avaliação de desmatamentos. Esse índice está sendo relacionado à cobertura vegetal do solo, biomassa e índice de área foliar, além de minimizar os efeitos de iluminação do local (Ponzoni & Shimabukuro, 2007).

O conhecimento do maior número possível de variáveis ambientais possibilita um melhor detalhamento de classes de solos. Com isso, torna-se importante conhecermos as variáveis que proporcionam mais influência sobre a variabilidade dos solos.

Quando temos um grande número de variáveis que podem se relacionar entre si, podemos utilizar a Análise de Componentes Principais (ACP), esta análise consiste na redução dos dados, formando um novo conjunto com a capacidade de manter ao máximo a variabilidade do conjunto inicial com perdas mínimas das informações (Regazzi, 2001). Essa técnica também permite agrupar as variáveis similares através da observação gráfica.

A seleção de variáveis que não tenham o comportamento de multicolinearidade dentro de variáveis ambientais também é feita através da aplicação de ACP (ten Caten et al., 2011a; 2011b).

Neste trabalho, objetivou-se conhecer as variáveis que mais influenciam na predição de classes de solos através da ACP, com o intuito de reduzir o número de variáveis predictoras que são aplicadas ao MDS.

MATERIAL E MÉTODOS

A área de estudo está localizada na sub-bacia do Rio Vacacaí-Mirim, na região do Rebordo do Planalto do Rio Grande do Sul, região de transição entre o Planalto e a Depressão Central. A sub-bacia possui área de 1.892 ha, com relevo variando de ondulado a montanhoso, com declividades de 13 a 45%, podendo em alguns locais ocorrer declividades superiores a 100 % e com isso apresenta uma grande variabilidade de solos como: Argissolo Bruno- Acinzentado, Argissolo Vermelho, Neossolo Litólico, Planossolo Háplico, Neossolo Regolítico, Neossolo Flúvico e Cambissolo Háplico.

A altitude varia entre 200 e 480 m. O clima predominante é classificado conforme Köppen como Cfa, caracterizado por ser subtropical úmido sem estiagem, com temperatura média anual de 19,2°C e precipitação pluviométrica média anual de 1.708 mm, geralmente com boa distribuição durante as estações do ano. A geologia varia, podendo ser de origem vulcânica e sedimentar.

Para o melhor entendimento da metodologia empregada foram subdivididas em três etapas: aquisição do modelo digital do terreno MDT (i), obtenção das variáveis ambientais (ii) e obtenção das correlações e componentes principais (iii).

Aquisição do MDT

O modelo digital do terreno (MDT), no caso, a variável ELEV (elevação), foi originada da interpolação das curvas de nível da carta

topográfica. O método de interpolação utilizado foi *Thin Plate Spline (TIN)* do programa SAGA-GIS. A resolução espacial do MDT foi de 30 metros.

Obtenção das variáveis ambientais

Após a aquisição do MDT foram obtidas as variáveis ambientais, também chamadas de atributos do terreno, tais como: curvatura planar (CPLAN), curvatura de perfil (CPRF), fator topográfico (LS), umidade do terreno (WET) e declividade do terreno (DECL) utilizando a ferramenta *Basic Terrain Analysis* do SAGA-GIS.

As variáveis distância do ponto cotado (DCOT) e distância que estamos de uma drenagem (HIDRO) foram obtidas da carta topográfica e convertidas em arquivo raster para poderem ser utilizadas no programa GRASS-GIS, o qual calcula esses atributos.

A variável incidência da radiação solar (ILUS) foi obtida a partir da simulação de iluminação na superfície do MDT. Para isso foram escolhidos os dias em que as estações do ano mudam, e as datas foram inseridas no software Heliotec disponível no site http://www.heliodon.com.br/calc_sol_1.html, no qual se obteve os valores de azimute e inclinação do sol, utilizando a ferramenta *Hillshading* do SAGA-GIS, o qual gerou quatro mapas de simulação da superfície, que foram somados originando o atributo ILUS.

A variável desvio padrão dos índices de vegetação por diferença normalizada (DPNDVI), foi gerada a partir das imagens Landsat 5, utilizando as datas de 21/01/2009, 02/09/2010, 08/11/2012 e 26/12/2012.

Obtenção das correlações e componentes principais

As componentes principais (CP's) foram obtidas de uma matriz com 1100 amostras, representando a população de dez variáveis, que com a aplicação da ACP foi reduzida em uma matriz quadrada de dimensão de 10x10.

Para a obtenção das CP's e das correlações entre as variáveis fez-se primeiramente a padronização da matriz de dados. Posteriormente, foram obtidas as CP's e a variabilidade para as dez CP's, em que os resultados estão apresentados nas **tabelas 1 e 2** e **figura 1**. As correlações e os CP's foram obtidos utilizando o software livre R (R Development Core Team, 2011).

RESULTADOS E DISCUSSÃO

A criação de um novo conjunto de dados (**Tabela 1**) através da ACP possibilitou potencializar a

variabilidade encontrada entre as variáveis ambientais estudadas neste trabalho.

Avaliando a porcentagem de variância e da variância acumulada (**Tabela 1**), pode-se notar que as maiores porcentagens foram encontradas nas duas primeiras CP's analisadas, totalizando uma variância acumulada de 53,18%. Observou-se também que os resultados de variabilidade acumulada maior que 80% ficaram distribuídos entre as cinco primeiras componentes. Esses resultados foram obtidos a partir da rotação de um conjunto com dez variáveis ambientais e resultados semelhantes foram encontrados por Hengl & Rossiter (2003), ten Caten et al.(2011a) e ten Caten (2011), que também utilizaram a ACP para diminuir os efeitos de multicolinearidade entre as variáveis ambientais avaliadas.

Tabela 1: Resultado da distribuição da variabilidade nas dez componentes principais.

Componente Principal (CP)	Autovalores	Variância (%)	Variância acumulada (%)
1	3,8090	38,090	38,09
2	1,5090	15,090	53,18
3	1,1040	11,040	64,22
4	0,9573	9,573	73,79
5	0,9188	9,188	82,98
6	0,6771	6,771	89,75
7	0,6447	6,447	96,20
8	0,3589	3,589	99,79
9	0,0208	0,208	100,00
10	0,0000	0,000	100,00

Em contraponto, as oito primeiras CP's perfazem 99,79 % da variância acumulada, sendo que as demais CP's apresentam maior grau de independência quando comparado com as variáveis originais, apresentando valores residuais de variância.

Por outro lado, as componentes 9 e 10 por apresentarem uma variabilidade muito baixa, com valores residuais, são menos relevantes para serem aplicadas em modelos preditivos de classes de solos.

A exclusão de variáveis, que em geral, apresentaram menor variância poderá acarretar em perda do potencial preditivo do modelo, tendo em vista que a análise de componentes principais tem como objetivo reduzir o número de variáveis, tornando sua estrutura perceptível e interpretável. Estudos realizados por ten Caten et al. (2011a) comprovam que o uso de um menor número de variáveis não compromete a acurácia do modelo de predição de solos, o que também vem a colaborar

com a geração de menores custos e tempo para a obtenção das variáveis.

Na **figura 1** pode-se observar que houve a formação de pares ortogonais formados respectivamente pelas variáveis CPLAN e DCOT, LS e HIDRO, ELEV e DECL, os quais foram apresentando valores aproximados e encontram-se em posição simétrica no diagrama de ordenação.

Com o auxílio da **figura 1** e da **tabela 2** pode-se afirmar que as variáveis que mais influenciaram na maior variância da primeira componente principal foram DECL, CPRF e LS, estando essas correlacionadas entre si, pois estão muito próximas dentro do diagrama de ordenação. Em sentido oposto observa-se que a variável WET se correlaciona positivamente com a CP1, contribuindo também para a maior porcentagem de variância da CP1 (38,09%), porém esta variável se correlaciona inversamente com as demais variáveis, formando ângulo próximo de 180°.

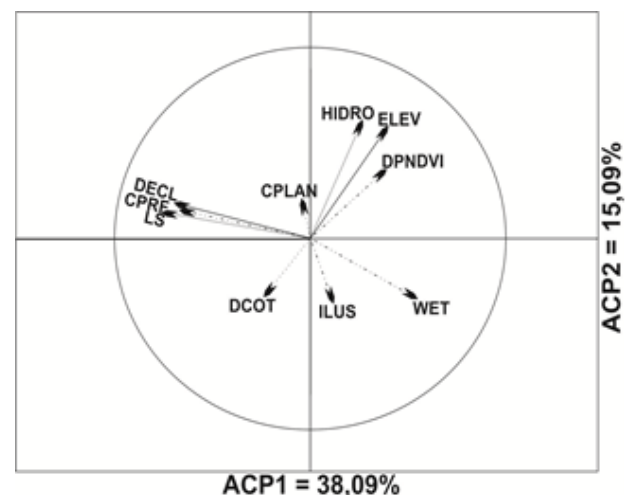


Figura 1- Diagrama de ordenação

Com uma relação praticamente simétrica às variáveis DECL, CPRF e LS, estão as variáveis HIDRO, ELEV e DPNDVI, que são correlacionados entre si em função da proximidade, conforme pode ser visto na **figura 1**. Além disso, essas variáveis contribuem para o segundo maior valor de variância (15,09%), correspondente à CP2. As variáveis DCOT, ILUS e WET também contribuem para a variância da CP2, estando em posições aproximadamente opostas aos outros três atributos, porém com uma menor contribuição na variabilidade desta componente.

Por outro lado, analisando a **tabela 2**, pode-se verificar que a variável DCOT tem maior influência sobre a CP4, bem como a variável ILUS, embora essa componente tenha menor variabilidade. A variável CPLAN não tem correlação com as demais

variáveis (**Figura 1**) e tem menor contribuição na variabilidade das componentes com maior valor de variância, influenciando positivamente e de forma considerável na variância da CP3 (**Tabela 2**).

Com isso, pode-se inferir que entre os dois grupos de variáveis que se correlacionam e influenciam nos maiores valores de variância, podemos excluir os atributos LS, DPNDVI e HIDRO. Essas variáveis não são consideradas boas preditoras de classes de solos, por motivo que a variável LS esta correlacionada com as variáveis DECL e CPRF, e a variável HIDRO estar correlacionada com as variáveis ELEV e DPNDVI (**Figura 1**), e esta última também apresentar menor representatividade de influência na variabilidade (**Tabela 2**).

Tabela 2: Matriz de correlação entre as cinco primeiras CP's

Atributos	PC1	PC2	PC3	PC4	PC5
CPRF	-0,4930	0,1550	-0,0808	0,0138	-0,1064
CPLAN	-0,0263	0,1642	0,8442	0,2509	0,0992
DCOT	-0,1419	-0,2714	0,3834	-0,5708	-0,4619
DECL	-0,4930	0,1551	-0,0808	0,0138	-0,1064
ELEV	0,2512	0,5183	-0,0359	-0,0285	-0,1238
HIDRO	0,1666	0,5465	-0,0734	0,1767	-0,1828
LS	-0,4734	0,1272	-0,2318	-0,0208	-0,1216
WET	0,3407	-0,2789	-0,2559	-0,1324	-0,0136
ILUSRP	0,0721	-0,3037	-0,0327	0,6373	-0,6929
DPNDVI	0,2384	0,3127	-0,0086	-0,3932	-0,4586

A variáveis ILUS também não seria significativa para a predição de classes de solos em função da que a mesma somente tem melhor influência na variabilidade da PC 4 (**Tabela 2**), e com isso também poderia ser excluída do conjunto de variáveis preditoras, concordando com Debella-Gilo & Etzelmüller (2009).

As variáveis CPLAN, CPRF, DECL, ELEV, WET e DCOT se comportam melhor como variáveis preditoras, tendo aptidão para constituir modelos de MDS para a área de treinamento.

Desta forma o conjunto de variáveis que antes era composto por dez variáveis passa a ser, após a ACP, reduzido para cinco, as quais poderão ser aplicadas em estudos sequencias de MDS.

CONCLUSÕES

A ACP permitiu reduzir e prever o conjunto de variáveis ambientais que não se correlacionam entre si e que seriam as variáveis preditoras, neste caso, de classes de solos do Rebordo do Planalto do RS.

AGRADECIMENTOS

Ao PPGCS da UFSM e à CAPES pela concessão de recursos financeiros e bolsa de estudo.

REFERÊNCIAS

- COELHO, F. F. & GIASSON, E. Comparação de métodos para mapeamento digital de solos com utilização de sistema de informação geográfica. *Ci. Rural*, v. 40; p.2099-2106, 2010.
- CORRÊA, M. P. Influência da radiação solar na regeneração natural de mata atlântica. 2008. 38f. Monografia (Curso Eng. Florestal), Instituto de Florestas - Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2008.
- 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.
- HENGL, T. & ROSSITER, D.G. Supervised Landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Sci. Soc. America Journal*. v.67, p.1810-1822, 2003.
- McBRATNEY, A. B.; MENDONÇA SANTOS, M. L.; MINASNY, B. On digital soil mapping. *Geoderma*, v.117, p.3-52, 2003.
- PEREZ, L. P et al. Dinâmica dos principais domínios fitogeográficos do nordeste brasileiro e suas conexões com a pluviometria. *Geografia*, v.29, 2004.
- PONZONI, F. J. & SHIMABUKURO, Y. E.; Sensoriamento Remoto no Estudo da Vegetação. Parentese: Sao Paulo, 2007.
- R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. 2011.
- REGAZZI, A. J. Análise multivariada. 2001. 166p. Apostila de disciplina. Centro de Ciências Exatas e Tecnológicas, Departamento de Informática, Universidade Federal de Viçosa. Viçosa, 2001.
- ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; Deering, D. W. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. *Prog. Rep. RSC. Remote Sensing Cent., College Station*, 1973.
- TEN CATEN, A. et al. Mapeamento Digital: probabilidade associada à distribuição espacial de classes de solos. *Geomática*, v.4, p.87-97, 2009.
- TEN CATEN, A. et al. Estatística multivariada aplicada à diminuição do número de preditores no mapeamento digital de solos, *Pesq. agropec. bras.*, v. 46, n.5, p.554-562, 2011a.
- TEN CATEN, A et al. Componentes principais como preditores no mapeamento digital de classes de solos, *Ci. Rural*, v.41, n.7, p.1170-1176, 2011b.
- TEN CATEN, A. Mapeamento digital de solos: Metodologias para atender a demanda por informação espacial em solos. 2011. 106f. Tese de Doutorado. (Ciência do Solo), Universidade Federal de Santa Maria, Santa Maria, 2011.