

## Mapeamento digital de solos através da aplicação de componentes principais em modelos logísticos.

Alexandre ten Caten<sup>1</sup>

Ricardo Simão Diniz Dalmolin<sup>2</sup>

Luiz Fernando Chimelo Ruiz<sup>3</sup>

Elódio Sebem<sup>4</sup>

Rudiney Soares Pereira<sup>5</sup>

<sup>1,2</sup>Universidade Federal de Santa Maria – Departamento de Solos  
Av. Roraima, 1000 – CCR II – cep: 97105/900 – Santa Maria / RS

<sup>1</sup>acaten@yahoo.com.br, <sup>2</sup>dalmolinrsd@gmail.com

<sup>3,4</sup>Colégio Politécnico da UFSM / Av. Roraima, 1000 cep: 97105/900 – Santa Maria / RS

<sup>3</sup>ruiz.ch@gmail.com, <sup>4</sup>elodiosebem@gmail.com

<sup>5</sup>Universidade Federal de Santa Maria – Departamento de Engenharia Rural

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

<sup>5</sup>rudiney.s.pereira@gmail.com

**Abstract:** Digital Soil Mapping (DSM) is an interdisciplinary science involving soil, statistics, mathematics and geomatics knowledge applied into generating soil spatial information. This study aimed to use Principal Component (PC) as covariates in logistic models for the prediction of soil classes in areas not mapped. Principal Component Analysis (PCA) was applied to nine terrain attributes: elevation, slope, distance to nearest stream, planar curvature, profile curvature, illumination, natural logarithm of the contributing area, topographic wetness index and sediment transport capacity. The retained PC were used as explanatory covariates in Multiple Logistic Regressions (MLR), which were trained with soil information provided by an available soil map on a 1:50.000 scale. The three retained components explained 65.57% of the variability on the original data. The terrain attributes significantly related to the first CP were distance to nearest stream, slope and elevation. The logistic models reproduce the original map in 58.20% (kappa index), and the predictive ability of models was 48.53%. Soil units with the smallest areas were not properly spatialized, and logistic models were not able to distinguish between soil classes too close on the landscape. Future studies have to improve the use of PCA with a greater number of covariates, searching for the applicability of the PC in MLR for soil class prediction.

**Palavras-chave:** Pedometria, índice de umidade topográfica, mapa de solos, SRTM, pedometry, topographic wetness index, soil map, SRTM.

### 1. Introdução

#### 1.1 Mapeamento Digital de Solos (MDS)

O MDS tem como objetivo a criação de sistemas de informações espaciais em solo, através do uso de métodos e observações em campo e laboratório, associados a sistemas de inferência espaciais e não espaciais (LAGACHERIE e MCBRATNEY, 2007). A disponibilização de tecnologias as quais geram dados sobre os fatores e processos condicionantes do solo, aliada ao emprego de técnicas matemáticas e estatísticas, possibilita que o MDS seja empregada para atender a uma demanda crescente por informação espacial em solos. O MDS tem se valido da disponibilização de tecnologias para a coleta de dados espaciais como o Sensoriamento Remoto por Radar. O Modelo Digital de Elevação (MDE) gerado a partir do *Shuttle Radar Topography Mission* (SRTM) possibilitou que uma série de estudos, especialmente considerando o fator “r” da equação de Jenny (1941), fossem realizados visando à predição espacial de classes, características e propriedades do solo.

## 1.2 Análise de Componentes Principais (ACP)

A ACP é um método multivariado que permite transformar um conjunto de variáveis iniciais correlacionadas entre si, num outro conjunto de variáveis não correlacionadas, as chamadas Componentes Principais (CP). É uma técnica poderosa que pode ser utilizada para a redução do número de variáveis e para fornecer uma visão privilegiada do conjunto de dados. Algebricamente, as componentes são combinações lineares ponderadas das  $p$  variáveis aleatórias  $X_1, X_2, \dots, X_p$ , equação 1 (JOHNSON e WICHERN, 1992).

Generalizando para  $m$  componentes e  $p$  variáveis ( $m \leq p$ ):

$$\begin{aligned} CP_1 &= a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ CP_2 &= a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ &\dots \\ CP_m &= a_{1m}X_1 + a_{2m}X_2 + \dots + a_{pm}X_p \end{aligned} \quad (1)$$

Onde,

$CP_m$ : m-ésima componente principal;

$a_{pm}$ : autovetor ou peso de cada uma das variáveis para a CP;

$X_p$ : variáveis observadas.

Para Webster (2001) a ACP constitui-se em uma transformação de dados, sendo uma rotação rígida dos dados originais para um novo conjunto de eixos. É um procedimento matemático, não estatístico. Não requer pressuposição da normalidade da distribuição e leva a nenhum teste estatístico de significância. A ACP, por estar baseada em um modelo linear, tem boa aplicabilidade aos estudos que relacionem o solo a preditores ambientais, pois raramente existe entre eles uma relação não linear (Gaussiana) (ODEH et al., 1991).

## 1.3 Regressões Logísticas Múltiplas (RLM)

É uma técnica muito mais flexível pois 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 (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 por:

$$\text{Logit}_i = \ln\left(\frac{P_i}{(1 - P_i)}\right) = a + b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (2)$$

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. Como é mais fácil de se perceber a razão entre probabilidades do que seu logaritmo, a equação 2 pode ser reescrita como:

$$\frac{P_i}{(1 - P_i)} = e^{a+b_0+b_1X_1+b_2X_2+\dots+b_kX_k} \quad (3)$$

Tratando-se de um conjunto de respostas politômicas (várias classes de solo), faz-se necessário estabelecer uma categoria de referência ou base, da qual, em um primeiro momento, não é calculada a probabilidade de ocorrência. Para as demais classes a probabilidade de ocorrência é dada por:

$$P_i = \frac{e^{a+b_0+b_1X_1+b_2X_2+\dots+b_kX_k}}{1 + \sum_1^{m-1} (e^{a+b_0+b_1X_1+b_2X_2+\dots+b_kX_k})} \quad (4)$$

A qual é usada para prever a probabilidade  $P_i$  de que a classe de solo “i” esteja presente em determinado pixel, dadas as variáveis preditoras  $X_1, X_2, \dots, X_k$ . Através da razão entre o logit de “i” e a soma dos logit de todas as outras classes, exceto da classe tomada como referência, mais um.

A classe tomada como referência, não terá seu logit determinado. Sendo sua probabilidade determinada por:

$$P_i = \frac{1}{1 + \sum_1^{m-1} (e^{a+b_0+b_1X_1+b_2X_2+\dots+b_kX_k})} \quad (5)$$

Todos os valores de “a” e “b” das regressões serão determinados para cada classe de solo a partir de dados experimentais (DEBELLA-GILO et al., 2007).

Esse estudo teve como objetivo utilizar as CP como parâmetros dos modelos logísticos 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. Possui uma superfície de 873,59 km<sup>2</sup>, e está compreendido entre as coordenadas 29°46' a 29°26' de latitude sul e 54°30' a 53°56' de longitude oeste. Abrange uma região transitória entre as regiões fisiográficas do Planalto Médio e Depressão Central. Esta área foi escolhida por apresentar um levantamento de solos semidetalhado na escala 1:50.000 (KLAMT et al., 2001).

### 2.2 Extração das componentes principais

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), foram gerados de acordo com o procedimento detalhado por Caten (2008) a partir de um MDE / SRTM. Um conjunto de 70.000 pontos foi aleatoriamente gerado para amostrar os Planos de Informação (PI) de atributos de terreno, estas informações tabuladas em formato texto (ASCII) foram processadas para a ACP. Tendo sido verificada a adequabilidade das amostras pelo teste individual *Measure of Sample Adequacy* (MAS) e geral *Kaiser Meyer Olkin* (KMO) segundo GOBIN et al. (2000), visando à verificação do grau de correlação entre as variáveis e a justificativa da ACP. O número de autovalores retidos foi condicionado para aqueles que tivessem valor mínimo igual a um. Os autovetores rotacionadas (VARIMAX), resultantes da ACP, foram utilizados para o cálculo dos valores das novas variáveis.

### 2.3 Mapa predito a partir do mapa de solos existente

As RLM foram geradas utilizando-se das CP como variáveis explicativas, e das classes de solos no mapa existente ao nível de ordem (1º 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. A qualidade dos mapas gerados foi

avaliada quanto a sua capacidade em reproduzir o mapa original, e quanto a capacidade em prever as classes de solos para uma área de onde não foram utilizados dados para a geração dos modelos.

### 3. Resultados e Discussão

#### 3.1 Componentes principais

Para a aplicação da análise de CP na matriz de correlação dos atributos, primeiramente foi verificada a adequabilidade dos dados pelo testes individual MSA e geral KMO. Como preconizado pela literatura, valores de MSA abaixo de 0,5 indicam que a variável não é apropriada para a aplicação da ACP. Entre os atributos do terreno CPRF e LNAC obtiveram valores de 0,58 e 0,56, respectivamente, os quais podem ser considerados valores bastante baixos para a aplicação destas variáveis em ACP. No entanto, como o número de atributos é de apenas nove, optou-se por permanecer com todas as variáveis. O valor de KMO do conjunto de todos os atributos foi de apenas 0,66, também podendo ser considerado baixo para aplicação da ACP, contudo ainda praticável.

Após a aplicação da ACP aos nove atributos do terreno foram gerados nove CP, cada uma concentrando um percentual decrescente da variabilidade dos dados iniciais (Figura1).

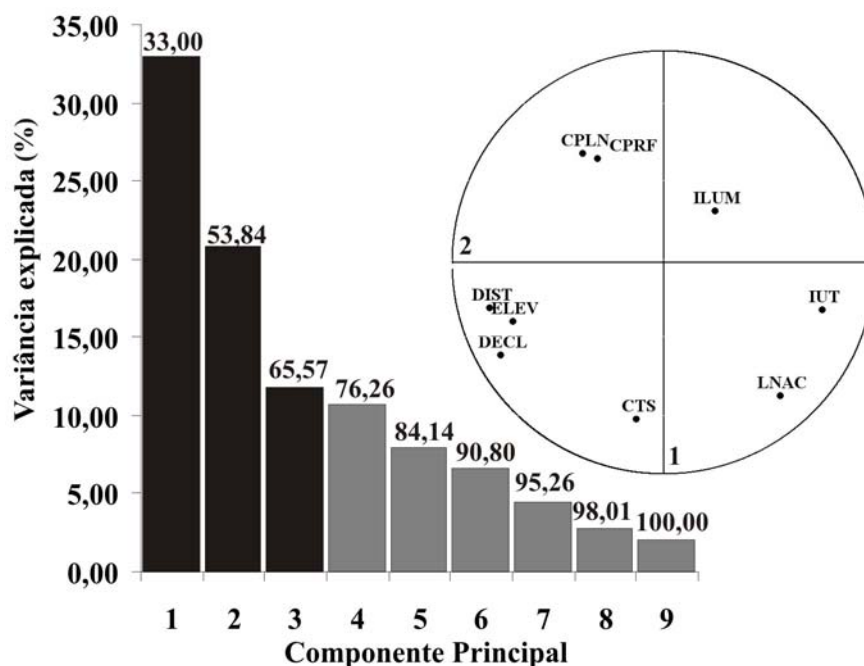


Figura 1 – Variância explicada por cada uma das nove componentes principais (CP) geradas. Valor sobre as barras indica o percentual de variância acumulada. Diagrama de dispersão circular indica as correlações entre variáveis e a primeira (1) e segunda (2) CP. CP com barras em cor preta foram utilizadas como variáveis explicativas neste estudo. 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).

As três primeiras CP possuem autovalores maior do que um e foram retidas nesse estudo. Quanto ao percentual de variância explicada pelas novas variáveis, verifica-se que a primeira componente é responsável por um terço da variância dos dados originais. A segunda e a terceira componentes têm juntas um poder de explicação equivalente ao da primeira componente, em torno de um terço da variabilidade total. A retenção apenas das três primeiras componentes, significa uma perda de um terço da variabilidade dos dados acumulada nas novas variáveis da quarta a nona componente. Ganha-se com a simplificação do número de

variáveis, embora havendo uma perda significativa do padrão de variabilidade dos dados. Hengl (2003) utilizou ACP para mitigar o efeito de multicolinearidade entre atributos do terreno. O autor identificou que as cinco primeiras componentes foram responsáveis por cerca de 80% da variabilidade de um conjunto de nove atributos do terreno. Este resultado é muito semelhante ao reportado pela figura 1, onde as cinco primeiras componentes acumulam cerca de 84,14% da variância.

Os pesos dos nove atributos do terreno nas duas primeiras componentes retidas podem ser visualizados no diagrama de ordenação unitário da figura 1. Os atributos mais significativos para os 33,0% da variabilidade explicada pelo primeiro eixo são DIST, DECL e ELEV, os quais são correlacionados entre si como indica a sua proximidade dentro do círculo unitário. Da mesma forma, o IUT tem correlação positiva com a primeira CP. As contribuições dos atributos do terreno na segunda componente, com uma variância de 22,84%, podem ser visualizadas nas projeções destes atributos para a linha vertical do círculo unitário. Os atributos mais significativos são a CTS e a LNAC, embora ainda com correlações bastante baixas com esta componente. Os valores de CPLN e CPRF mostraram estas variáveis bastante correlacionadas entre si nestas duas primeiras componentes.

Estudos como os realizados por Debella-Gilo et al. (2007) e Giasson et al. (2008) apontaram que os atributos do terreno como ELEV e DECL tiveram grande importância para os modelos logísticos, no entanto, os autores não fazem nenhuma menção quanto a multicolinearidade das variáveis. Os dados gerados neste estudo e analisados por CP demonstram que estas variáveis foram correlacionadas.

### 3.1 Mapa de solos predito

A tabela 2 traz os valores de intercepto e coeficientes dos modelos logísticos.

Tabela 2 – Valores dos interceptos e dos coeficientes das componentes principais nos modelos logísticos desenvolvidos ao nível de ordem.

	C <sup>1</sup>				
	R	SH	P	N	F
<b>Intercepto</b>	-1,168	-7,805	-0,562	-2,612	-8,104
<b>CP1</b>	-0,002	-0,115	-0,058	-0,013	-0,127
<b>CP2</b>	-0,005	0,042	-0,032	0,005	-0,056
<b>CP3</b>	0,002	0,288	0,127	0,032	0,287

<sup>1</sup>Classe de referência. Todos os valores da tabela 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 (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N), Plintossolo (F), Primeira Componente Principal (CP1), Segunda Componente Principal (CP2) e Terceira Componente Principal (CP3).

O mapa de solos gerado não espacializou as classes de solos Neossolos, Plintossolos e Nitossolos. Embora as classes de solos Cambissolos, Argissolos e os Solos Hidromórficos tenham sido espacializadas visualmente de maneira muito semelhante ao que se encontra no mapa existente, e na relação solo-paisagem da área de estudo. Os Solos Hidromórficos foram espacializados nas partes mais baixas da paisagem e os Cambissolos nas encostas e topos de morros. Ao passo que os Argissolos foram distribuídos nas coxilhas, e a classe dos Cambissolos foi atribuída a regiões onde encontram-se os Neossolos. Supõe-se que a motivação para a não espacialização das classes Neossolos, Plintossolos e Nitossolos tenha sido devido a pouca representatividade destas classes no total das amostras utilizadas nos

modelos logísticos. Estas classes correspondem somente a cinco, três e um % do total das 70.000 amostras randomicamente selecionadas para a geração dos modelos.

Para melhorar a capacidade dos modelos 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) afirma 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.

A análise da tabela 3 permite verificar que a acurácia geral (AG) do mapa predito foi de 74,3%, este valor pode ser considerado bom, embora não considere os pontos corretamente mapeados devido ao acaso. Para contornar isso se considera o valor do índice kappa (K) de 58,20% como uma medida mais realista da qualidade do mapa predito.

Tabela 3 – Matriz de erros da área externa, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir das componentes principais como variáveis dos modelos logísticos.

Mapa predito	Mapa Klamt						Total	AU (%)
	R	SH	P	N	F	C		
R	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SH	0,03	17,92	4,69	0,01	1,55	0,12	24,33	73,65
P	0,86	4,93	39,72	0,57	1,15	4,42	51,66	76,89
N	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C	4,68	0,12	2,28	0,19	0,09	16,66	24,02	69,36
<b>Total</b>	5,57	22,97	46,69	0,77	2,79	21,20	100,00	
<b>AM (%)</b>	0,00	78,01	85,07	0,00	0,00	78,58		
<b>AG (%)</b>	74,30		K <sup>1</sup> (%)		58,20			

<sup>1</sup>Erro padrão assintótico = ± 0,009 e intervalo de confiança 95% = 0,56...0,60. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH), Argissolo (P), Nitossolo (N) e Plintossolo (F).

Os erros mais comuns na matriz foram entre classes espacialmente próximas quanto ao delineamento no mapa. Os solos de várzea foram confundidos com os Argissolos, estes por sua vez foram erroneamente espacializados nas posições dos Cambissolos, os quais confundiram-se com os Neossolos. Estes erros da matriz podem ter origem nas bordas de cada classe de solos. A inferência da verdadeira classe, sob o ponto de vista do mapa existente, pode ser bastante difícil para os modelos devido a problemas no próprio delineamento que serviu de treinamento. Ou ainda, devido a diferenças muito tênues entre os atributos do terreno, os quais podem não apresentar nenhum tipo de gradiente na borda dos polígonos.

Boruvka e Penizek (2007) utilizaram redes neurais para a espacialização de classes de solos e verificaram que classes muito semelhantes sob o ponto de vista dos processos de formação tendem a ser confundidas pelos modelos. As classes devem ser bem definidas e distintas entre elas para uma boa espacialização, o emprego de qualquer metodologia, segundo os autores, deverá considerar o nível categórico a ser predito em função da heterogeneidade local, além a disponibilidade de informações para a geração dos modelos (número de perfis ou área).

Para a verificação da capacidade preditiva do modelo utilizado, foi realizado um teste de acurácia em uma área da qual não foram utilizados dados para a geração do modelo. Este procedimento valida o modelo utilizado, ou de outra maneira, testa a sua real capacidade de inferência ou capacidade preditiva, os resultados estão na tabela 4. A AG chegou a ser até

mesmo maior do que a AG da área de onde foram usados pontos para o modelo, alcançando um valor de 79,4%. No entanto, uma medida mais realista da qualidade do mapa nesta região, mostra que a acurácia do mapa foi menor, o índice K de 48,53% ficou dez pontos percentuais a baixo da área de onde foram utilizados pontos para o modelo.

Tabela 4 – Matriz de erros da área interna, acurácia do mapeador (AM), acurácia do usuário (AU), acurácia geral (AG) e índice kappa (K) entre o mapa Klamt e o mapa predito ao nível de ordem produzido a partir das componentes principais como variáveis dos modelos logísticos.

Mapa predito	Mapa Klamt				Total	AU (%)
	R	SH	P	C		
R	0,00	0,00	0,00	0,00	0,00	0,00
SH	0,00	6,50	4,30	0,00	10,80	60,19
P	0,20	8,40	67,10	1,90	77,60	86,47
C	2,40	0,00	3,40	5,80	11,60	50,00
<b>Total</b>	2,60	14,90	74,80	7,70	100,00	
<b>AM (%)</b>	0,00	43,62	89,71	75,32		
<b>AG (%)</b>	79,40		<b>K<sup>1</sup> (%)</b>	48,53		

<sup>1</sup>Erro padrão assintótico =  $\pm 0,035$  e intervalo de confiança 95% = 0,41...0,55. Cambissolo (C), Neossolo (R), Solos Hidromórficos (SH) e Argissolo (P).

Os resultados desse estudo demonstram o potencial da ACP para redução do número de variáveis a serem aplicados em modelos, além de permitir a visualização das correlações entre as variáveis originais e produzir novas variáveis não correlacionadas. Embora a aplicação de ACP implica em perda da variabilidade original dos dados, o uso das três primeiras CP permitiu que os modelos logísticos espacializassem as classe de solos com uma acurácia, em relação a mapa original, de em torno de 60% o que é compatível com dados da literatura. O uso da ACP provavelmente será potencializado com um maior número de variáveis predictoras, e com melhores valores de MSA e KMO.

As RLM podem ser mais efetivas para a espacialização de classes de solos se estas classes tiverem uma representatividade relativa maior entre elas. Corroborando com dados da literatura, esse estudo demonstrou que classes de solos com pouca área na paisagem não são adequadamente espacializadas pelos modelos logísticos.

#### 4. Conclusões

A utilização de componentes principais em regressões logísticas múltiplas implica em modelos simplificados, comparativamente a utilização de todas as variáveis originais nos modelos. No entanto, essa simplificação poderá vir associada à uma redução do poder de explicação das novas variáveis devido a menor variância retida, além ainda, das novas variáveis não terem nenhum significado físico, químico ou biológico para a formação do solo.

Regressões logísticas múltiplas são efetivas na espacialização de classes de solos a partir de mapas existentes para áreas não mapeadas. Mostrando-se como uma ferramenta que precisa ser melhor estudada para gerar informações em solos ainda tão necessárias em nosso país.

Os modelos desempenharam com uma qualidade inferior quando utilizados fora da área de treinamento, sendo assim, a capacidade preditiva dos modelos estará associada a geração dos modelos em áreas similares, do ponto de vista de processos e fatores de formação do solo, àquelas onde os modelos serão aplicados.

## 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.

BORUVKA, L.; PENIZEK, V. A test of an artificial neural network allocation procedure using the Czech Soil Survey of Agricultural Land data. . In: Lagacherie, P.; Mcbratney, A.; Voltz, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 415-424.

CATEN, A. ten. **Aplicação de componentes principais e regressões logísticas múltiplas em sistema de informações geográficas para a predição e o mapeamento digital de solos**. 2008. 128f. Dissertação (Mestrado em Ciência do Solo) – Universidade Federal de Santa Maria, Santa Maria, 2008.

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

DEBELLA-GILO, M.; ETZELMULLER, B.; KLAKEGG, O. Digital soil mapping using digital terrain analysis and statistical modeling integrated into GIS: Examples from Vestfold county of Norway. ScanGIS'2007. In: Scandinavian Research Conference on Geographical Information Sciences, 11., 2007, Norway. **Proceedings...** Norway, 2007, 274 p.

GIASSON, E.; FIGUEIREDO, S.R.; TONRQUIST, C.G.; CLARKE, R.T. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in southern Brazil. In: Hartemink, A. E.; Mcbratney, A.; Mendonça Santos, M. DE L. (Ed.). **Digital soil mapping with limited data**. New York: Springer-Verlag, 2008, p. 225-232.

GOBIN, A.; CAMPLING, P.; DECKERS, J.; FEYEN, J. Quantifying Soil Morphology in Tropical Environments: Methods and Application in Soil Classification. **Soil Science Society American Journal**, v. 64, p. 1423-1433, 2000.

HENGL, T. **Pedometric mapping: bridging the gaps between conventional and pedometric approaches**, 2003, 236 f. Tese (Doutorado). Wageningen University, Enschede.

JENNY, H. **Factors of soil formation: a system of quantitative pedology**. New York, London: McGraw-Hill, 1941, 324 p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall, 1992. 642 p.

KLAMT, E.; FLORES, C.A.; CABRAL, D.R. **Solos do Município de São Pedro do Sul**. Departamento. de Solos/CCR/UFSM. Santa Maria, 2001, 96 p.

LAGACHERIE, P.; MCBRATNEY, A. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: Lagacherie, P.; Mcbratney, A.; Voltz, M. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007, p. 3-22.

ODEH, I.O.A.; CHITTLEBOROUGH, D.J.; MCBRATNEY, A. Elucidation of soil-landform interrelationships by canonical ordination analysis. **Geoderma**, Amsterdam, v. 49, p. 1-32, 1991.

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.

WEBSTER, R. Statistics to support soil research and their presentation. **European Journal of Soil Science**, v. 52, p. 331-340, 2001.