



XXXIII Congresso Brasileiro de Ciência do Solo

Solos nos biomas brasileiros: sustentabilidade e mudanças climáticas
31 de julho à 05 de agosto - Center Convention - Uberlândia/Minas Gerais

CLASSIFICAÇÃO SUPERVISIONADA DE SOLOS POR REDES NEURAIS ARTIFICIAIS NA SERRA DO CIPÓ - MG

Eliana de Souza⁽¹⁾; Elpídio Inácio Fernandes Filho⁽²⁾; César da Silva Chagas⁽³⁾; Carlos Ernesto G. R. Schaefer⁽²⁾; João Carlos Ker⁽²⁾; Carlos Antônio Oliveira Vieira⁽⁴⁾; Felipe Nogueira Bello Simas⁽²⁾

⁽¹⁾ Geógrafa, Estudante de doutorado em Solos e Nutrição de Plantas, Departamento de Solos (DPS), Universidade Federal de Viçosa (UFV). Av. P.H. Rolfs s/n. Campus Universitário. CEP 36571 000. Viçosa, MG. elianadsouza@yahoo.com.br; ⁽²⁾ Professor do DPS/UFV; ⁽³⁾ Pesquisador da Embrapa Solos. Rua Jardim Botânico 1024, Jardim Botânico, CEP 22460-000, Rio Janeiro, RJ; ⁽⁴⁾ Professor do DPS/UFV; ⁽⁵⁾ Professor do Departamento de Geociências Universidade Federal de Santa Catarina, Florianópolis, SC

Resumo – A classificação supervisionada de solos, especialmente nas últimas décadas, vem sendo realizada por modelos, matemáticos e estatísticos a exemplo das redes neurais. A maioria dos trabalhos com redes neurais avaliaram propriedades dos solos, sendo o estudo das classes ainda incipiente. Assim este trabalho teve como objetivo avaliar a classificação de solos por redes neurais para uma área situada na Serra do Cipó. Para tanto utilizou informações analíticas de perfis de solos e variáveis discriminantes como: bandas de imagem de satélite e índices derivados, mapas geológico, pedológico e de atributos do terreno. Os resultados obtidos na classificação pelas redes neurais foram comparados com uma classificação feita pelo classificador MaxVer a partir de validação dos mapas com pontos de referência terrestre. Os mesmos pontos foram utilizados para validar o mapa obtido pelo método convencional. Os mapas obtidos pelos dois classificadores apresentaram índice de exatidão considerado bom, sem diferença estatística. O mapa melhor classificado pelo MaxVer apresentou índice kappa de 0,58, enquanto que pelas redes neurais o maior índice foi de 0,60. A eficiência dos classificadores, na discriminação individual das unidades de solo se diferiram, sendo duas unidades melhor classificadas pelo MaxVer, três pelas redes neurais e quatro unidades com exatidão estatisticamente igual para os dois classificadores. A exatidão global do mapa de solos obtido pelo método convencional foi de 82%, obtido pela concordância dos solos de referência com aqueles associados em qualquer componente da unidade. Os solos no primeiro componente das unidades de mapeamento apresentaram 48% de concordância com solos de referência.

Palavras-Chave: Mapeamento de Solos, mapeamento digital, Redes Neurais, classificação de Solos, Serra do Espinhaço.

INTRODUÇÃO

Atualmente, os mapas de solos produzidos por meio do levantamento convencional são as principais fontes de informação pedológica para modelagem e atividades de manejo, contudo há uma carência de mapas em media escala. As abordagens de

mapeamento baseadas em modelos quantitativos são conhecidas na ciência do solo como pedometria e são utilizados para descrever, classificar e estudar padrões de distribuição espacial de propriedades e classes de solos, a partir de observações de campo e correlação de solos com variáveis ambientais (McBratney et al., 2003; Rossiter, 2005).

As pesquisas para o desenvolvimento dos modelos de predição dos solos têm sido focadas em estudos dos dados utilizados, em sistemas de predição da distribuição espacial dos solos e em métodos para mensurar a qualidade das informações classificadas (Hempel et al., 2008).

Neste contexto o presente trabalho teve os seguintes objetivos: 1) Mapear os solos em escala de reconhecimento de alta intensidade, com o auxílio dos classificadores supervisionados de RNA e Máxima Verossimilhança (MaxVer); 2) Determinar as variáveis preditivas que melhor discriminam os solos nos modelos de classificação por RNA e MaxVer e 3) Avaliar a eficiência dos classificadores de RNAs e do MaxVer na predição dos solo por meio de índices de exatidão calculado com pontos de referência terrestre e 4) Avaliar a exatidão de um mapa em escala de 1:100.000 feito pelo método convencional de mapeamento a partir de pontos de referência.

MATERIAL E MÉTODOS

A área de estudo abrange parte do Parque Nacional da Serra do Cipó e a Área de Proteção Ambiental Morro da Pedreira, situadas na Serra do Cipó, borda oeste da Cordilheira do Espinhaço, no Estado de Minas Gerais.

O mapeamento dos solos seguiu as normas estabelecidas para o levantamento pedológico na escala de 1:100.000, conforme EMBRAPA (1995). As informações pedológicas e constou de mapa em escala de 1:100.000 (Schaefer et al., 2008) e 55 perfis que representaram classificadas até o quarto nível categórico do Sistema Brasileiro de Classificação. A área foi percorrida para a aquisição 362 de pontos de referencia de solos para validar o mapeamento.

As variáveis discriminantes foram: seis cenas da imagem do satélite Landsat, sensor ETM+, quatro índices derivados dessa imagem (Clay minerals, Ferrous minerals, Iron oxide e NDVI); modelo digital de elevação e atributos derivados deste: altitude, declividade, índice topográfico combinado, face de exposição, radiação solar, curvatura e

amplitude altimétrica e mapas geológico e pedológico. A partir desse conjunto de 19 variáveis identificou aquelas que melhor contribuíram na discriminação dos solos em cada uma das duas abordagens empregadas.

Na classificação pelas redes neurais, empregou o simulador Stuttgart Neural Network Simulator e o algoritmo backpropagation, a arquitetura e parâmetros foram selecionados por meio de treinamentos e testes de significância estatística.

A validação da classificação foi feita utilizando a técnica multivariada por meio da matriz de confusão, a qual tem como indicadores de exatidão, o índice kappa, a exatidão global conforme indicado por Congalton e Green (1999). A concordância entre o mapa, bem como entre as unidades de solo obtidos na classificação pelas redes neurais e pelo MaxVer foi comparada empregando o teste Z.

RESULTADOS E DISCUSSÃO

A classificação foi feita inicialmente considerando cada uma das 20 classes de solo mapeadas, como uma unidade. A partir dos resultados da validação com amostras de treinamento foram observados os confundimentos entre as classes, e, redefinidas novas unidades, que na sequência foram reclassificadas. Esse processo foi realizado sucessivamente até considerar satisfatório o agrupamento dos solos em nove unidades, sendo três unidades simples, constituídas por uma classe cada uma e, seis unidades compostas por duas ou mais classes (Tabela 1).

A avaliação dos parâmetros da classificação pelas redes neurais, número de ciclos de treinamento e número de neurônio na camada interna das redes, por meio de testes estatísticos mostrou que a arquitetura com nove neurônios e 10 mil ciclos de treinamento apresentou apropriada para a classificação dos solos com as variáveis utilizadas.

A contribuição das variáveis discriminantes, na classificação pelas neurais e pelo MaxVer foi avaliada de forma interativa, removendo uma variável por vez do conjunto total da análise. Na classificação pelas redes neurais as variáveis que contribuíram em ordem crescente de importância para discriminar os solos foram: radiação solar, declividade, geologia e o índice Iron oxide. Por outro lado, o mapa de solos, o MDE, as bandas da imagem de satélite e o índice *Fe minerals* não apresentaram diferença estatística, mas contribuíram para aumentar a exatidão. Já o índice NDVI não alterou a exatidão do mapa, no entanto, com o uso do índice ocorreu maior exatidão para as unidades de Argissolos e Latossolos em detrimento dos Neossolos Litólicos. A variável aspecto, do mesmo modo, contribuiu para discriminar com maior exatidão determinadas unidades. O índice Clay mineral, por sua vez, foi a única variável que diminuiu a exatidão da classificação.

O classificador obteve o melhor desempenho quando utilizou o conjunto com 18 variáveis, sem o índice *Clay minerals*. A validação desse conjunto de variáveis com amostra de treinamento apresentou exatidão de 0,96 pelo índice kappa, sem diferença estatística em relação à exatidão da classificação com o

conjunto que incluiu todas as 19 variáveis, cujo índice foi 0,95.

Na classificação pelo MaxVer, as variáveis categóricas, geologia e curvatura não apresentaram contribuição para discriminar os solos, concorrendo com os maiores obtidos pelo classificador. A melhor classificação foi obtida quando essas duas variáveis e as bandas da imagem foram eliminadas do conjunto total de variáveis. Para esse conjunto foi gerado o mapa.

A validação dos mapas com amostra informacional de treinamento mostrou haver diferença significativa entre as classificações pelas redes neurais e pelo MaxVer. Entretanto na validação com pontos de referência terrestre todos os mapas apresentaram exatidão estatisticamente igual. Os mapas melhor classificados pelas duas abordagens apresentaram índice kappa de 0,60 pelas redes neurais e 0,58 pelo MaxVer. A exatidão foi maior para a classificação pelas redes neurais com 18 variáveis, apresentando 60% dos pontos corretamente classificados (kappa = 0,564). O MaxVer apresentou índices inferior ao obtido para as redes com 18 variáveis e superior ao das redes com 19 variáveis. A exatidão global do MaxVer foi de 59% e o índice kappa 0,547. A análise de significância estatística pelo Teste Z mostraram que os três mapas não apresentaram diferença estatística significativa na exatidão a 95% de probabilidade.

A avaliação da exatidão individual das unidades de solos classificadas pelas Redes Neurais e pelo MaxVer com base nos valores do índice kappa e a variância de cada uma das unidades (Tabela 2) foi realizada por meio do Teste Z. Esse teste comparou os resultados da classificação pelo MaxVer e pelas redes neurais com 18 e 19 variáveis.

Os valores de significância do teste Z (Tabela 3) em conjunto com os valores do índice kappa e variância obtidos para cada uma das unidades de solo mostram que: o MaxVer classificou duas unidades com exatidão estatisticamente superior a obtida na classificação pelas redes neurais (RRL + RRd e AR + RLd). Já as redes neurais classificou quatro unidades de solo com exatidão estatisticamente superior a apresentada pelo MaxVer (CXbd, RY, PVA, LVd + LVA). Três unidades não tiveram diferença estatística significativa entre a classificação pelo MaxVer e pelas redes neurais (PVd + PVAd, CXbd1, LVd + PVd).

A exatidão obtida para a classificação feita pelas redes neurais com os dois conjuntos de dados, 18 e 19 variáveis, mostrou diferença estatística significativa, para cinco das nove unidades mapeadas, com resultados superior para a classificação com o conjunto composto por 18 variáveis (CXbd1, LVd + LVA, RLd + RRd e RY). As demais quatro unidades não tiveram diferença estatística entre os resultados obtidos pelos dois conjuntos (PVAd, PVd + PVAd, CXbd, LVd + PVd e AR + RLd).

A concordância das classes de solo do mapa gerado pelo método convencional com pontos de referência terrestre foi calculada para cada um dos níveis de associação das unidades mapeamento. A exatidão global, do mapa, foi calculada a partir do somatório dos pontos concordantes em qualquer componente das unidades. A classificação dos solos para fins de validação foi considerada até o nível de subordens.

Os solos do primeiro componente de associação das unidades apresentaram 47,4%, de exatidão, 19,6% foi apresentada pelos solos do segundo e 14,9% pelos solos do terceiro. A exatidão global do mapa teve índice de exatidão de 81,8%, obtido pela concordância de todos os pontos corretamente classificados em qualquer um dos níveis das associações das unidades.

A concordância dos solos do primeiro componente teve o maior índice apresentado pelos Argissolos Vermelho-Amarelos com 84% de acerto, enquanto o menor índice ocorreu para os Latossolos Vermelhos, que obtiveram 19% de acerto. As demais classes apresentaram acerto entre 44 a 64%.

CONCLUSÕES

1. A avaliação das variáveis na discriminação dos solos em cada uma das abordagens empregadas foi importante para seleção daquelas que contribuíram para que os classificadores apresentassem maior exatidão no mapeamento.

2. Os classificadores MaxVer e Redes Neurais apresentaram diferença estatística significativa nos resultados do treinamento. No entanto, na validação com pontos de referência terrestre os resultados não se diferiram, mostrando importância de avaliar o treinamento e o mapa gerado de referência.

3. A exatidão dos mapas foi estatisticamente igual para as duas abordagens, embora os classificadores tenham apresentado diferença na discriminação das unidades, com a maioria das unidades melhor classificadas pelas redes neurais.

4. O mapa de solos obtido pelo método convencional apresentou os maiores acertos para os solos do primeiro componente das unidades.

AGRADECIMENTOS

Os autores agradecem à FAPEMIG pelo recurso concedido para realização dessa pesquisa.

REFERÊNCIAS

- CONGALTON, R. G. & GREEN, K. Basic Analysis Techniques, In: CONGALTON, R. G. & GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publishers. 1999.
- EMBRAPA. Serviço Nacional de Levantamento e Conservação de Solos. **Manual de métodos de análise de solo**. Rio de Janeiro. v1. 1979.
- HEMPEL, J.W. et al. Challenges to Digital Soil Mapping. In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital Soil Mapping with Limited Data**. Springer: Netherlands. p.60-81. 2008.
- McBRATNEY, A. B. et al. On digital soil mapping. **Geoderma**. v.117, p.3-52. 2003.
- ROSSITER, D. G. Digital soil mapping: towards a multiple-use Soil Information System. **Análisis Geográficos**. Augustin Codazzi. n.32, v.1. 2005.
- SANTOS, R. D. et al. **Manual de descrição e coleta de solo no campo**. 5. Ed. Viçosa. Sociedade Brasileira de Ciência do Solo/Embrapa-Centro Nacional de Pesquisa de Solos. 2005. 92p.
- SCHAEFER, C. E. R. G. et al. **Diagnóstico do meio físico da APA do Morro da Pedreira e Serra do Cipó subsídios ao plano de manejo: relatório do levantamento pedológico, geomorfológico e geológico**. IBAMA: 2008. 80 p.

Tabela 1 – Características das variáveis para as unidades de solo

Unidade de Solo Vegetação\ Relevo\ Geologia
PVAd - Argissolo Vermelho-Amarelo Distrófico típico, textura média\argilosa, A moderado. Floresta Semidecidual, relevo suave ondulado\ondulado\ Gnaisse.
PVd + PVAd - Argissolo Vermelho Distrófico típico, textura argilosa, A moderado; Argissolo Vermelho-Amarelo típico, textura média\argilosa, A moderado. Floresta Semidecidual, relevo ondulado\forte ondulado\Calcário.
CXbd - Cambissolo Háplico Tb latossólico, textura média e argilosa, A moderado; Cambissolo HáplicoTb Distrófico típico, textura média e argilosa, A moderado e proeminente. Floresta Estacional Nebular, relevo suave ondulado\forte ondulado; Floresta Estacional Nebular, relevo suave ondulado\ Filito.
CXbd1 - Cambissolo Háplico Tb Distrófico típico, textura média cascalhenta, A moderado, fase pedregosa. Cerrado, relevo ondulado\forte ondulado\ Siltito.
LVd + LVA - Latossolo Vermelho Distrófico típico, A moderado, textura argilosa; Latossolo Vermelho-Amarelo Distrófico típico, A moderado, textura argilosa. Relevo plano e suave ondulado; Cerrado, Floresta Decidual, relevo plano e suave ondulado\ Calcário.
LVd + PVd - Latossolo Vermelho Distrófico típico, A moderado, textura argilosa; Argissolo Vermelho típico, textura argilosa, A moderado. Floresta Semidecidual, relevo ondulado; Floresta Semidecidual, relevo ondulado\forte ondulado\ Gnaisse, Calcário.
RYd - Neossolo Litólico Distrófico psamítico, textura arenosa, A fraco e proeminente. Floresta ciliar, Relevo plano\ Sedimentos.
RLd + RRd - Neossolo Litólico Distrófico psamítico; Neossolo Regolítico Distrófico típico, textura arenosa, moderada e proeminente. Relevo plano a ondulado\ Quartzito, Filito.
AR + RLd - Afloramento de Rocha + Neossolo Litólico Tb Distrófico. Relevo montanhoso e escarpado\ Quartzito.

Tabela 2 – Valores obtidos na validação com pontos de referência terrestre para as classificações feitas pelo MaxVer e Redes Neurais

Unidade de solo	Kappa	Vâr.	Z _{calc}	Kappa	Vâr.	Z _{calc}	Kappa	Vâr.	Z _{calc}
PVAd	0,59	0,007286	9,0	0,62	0,009267	8,2	0,50	0,008147	7,8
PVd + PVAd	0,32	0,023093	3,7	0,55	0,031891	4,1	0,64	0,023794	5,2
CXbd1	0,25	0,003296	8,7	0,21	0,004596	6,8	0,28	0,002557	10,4
CXbd	0,59	0,019833	5,4	0,67	0,022893	5,4	0,37	0,018571	4,5
LVd + PVd	0,64	0,008369	8,7	0,69	0,010537	8,1	0,75	0,008461	9,4
LVd + LVA	0,36	0,007088	7,1	0,33	0,009508	5,9	0,18	0,006801	5,2
RLd + RRd	0,50	0,003622	11,7	0,14	0,003335	6,6	0,62	0,004286	12,0
RYd	0,76	0,014725	7,2	0,76	0,018390	6,4	0,15	0,009646	3,9
AR + RLd	0,82	0,009244	9,4	0,73	0,016919	6,6	1,00	0,000000	0,0
Mapa	kappa	Variância	Z	kappa	Variância	Z	kappa	Variância	Z
	0,56	0,000898	25,1	0,51	0,00092	23,5	0,55	0,000869	25,1

Conjunto de variáveis utilizado nas classificações: Redes Neurais 19 variáveis (todas as 19 variáveis); Redes Neurais 18 (eliminação da variável *clay mineral* do conjunto de 19 variáveis); MaxVer conjunto com 11 variáveis, (eliminação da geologia, curvatura e das seis bandas da imagem de satélite do conjunto de 19).

Tabela 3 – Matriz de significância estatística para as unidades de solo classificadas pelas Redes Neurais e pelo MaxVer

	PVAd			PVd + PVAd			CXbd1		
	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer
Redes (18)	9,0			3,7			8,7		
Redes (19)	0,2	8,2		1,4	4,1		10,9*	6,8	
MaxVer	5,5*	11,3*	7,8	4,4*	0,5	5,2	7,2*	1,6	10,4
	CXbd			LVd + PVd			LVd + LVA		
	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer
Redes (18)	5,4			8,7			7,1		
Redes (19)	0,5	5,4		0,5	8,1		9,2*	5,9	
MaxVer	2,9*	7,0*	4,5	7,0*	0,5	9,4	3,1*	7,7*	5,2
	RLd + RRd			RYd			AR + RLd		
	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer	Redes (18)	Redes (19)	MaxVer
Redes (18)	7,2			7,2			9,4		
Redes (19)	9,6*	6,4		9,6*	6,4		11,0*	6,6	
MaxVer	5,3*	10,7*	12,0	1,4	7,3*	3,9	11,5*	2,4*	0,0

* Existe diferença significativa a 5% de significância. Valor crítico de Z: 1,96. Os valores da diagonal principal da matriz representam a exatidão do classificador comparado a uma classificação aleatória, enquanto os valores de fora da diagonal representam o desempenho relativo entre os classificadores.