

INTEGRAÇÃO DE DADOS DO QUICKBIRD E ATRIBUTOS DO TERRENO NO MAPEAMENTO DIGITAL DE SOLOS POR REDES NEURAIAS ARTIFICIAIS⁽¹⁾

César da Silva Chagas⁽²⁾, Waldir de Carvalho Júnior⁽²⁾ & Silvio Barge Bhering⁽²⁾

RESUMO

No presente estudo, foi realizada uma avaliação de diferentes variáveis ambientais no mapeamento digital de solos em uma região no norte do Estado de Minas Gerais, utilizando redes neurais artificiais (RNA). Os atributos do terreno declividade e índice topográfico combinado (CTI), derivados de um modelo digital de elevação, três bandas do sensor Quickbird e um mapa de litologia foram combinados, e a importância de cada variável para discriminação das unidades de mapeamento foi avaliada. O simulador de redes neurais utilizado foi o “Java Neural Network Simulator”, e o algoritmo de aprendizado, o “backpropagation”. Para cada conjunto testado, foi selecionada uma RNA para a predição das unidades de mapeamento; os mapas gerados por esses conjuntos foram comparados com um mapa de solos produzido com o método convencional, para determinação da concordância entre as classificações. Essa comparação mostrou que o mapa produzido com o uso de todas as variáveis ambientais (declividade, índice CTI, bandas 1, 2 e 3 do Quickbird e litologia) obteve desempenho superior (67,4 % de concordância) ao dos mapas produzidos pelos demais conjuntos de variáveis. Das variáveis utilizadas, a declividade foi a que contribuiu com maior peso, pois, quando suprimida da análise, os resultados da concordância foram os mais baixos (33,7 %). Os resultados demonstraram que a abordagem utilizada pode contribuir para superar alguns dos problemas do mapeamento de solos no Brasil, especialmente em escalas maiores que 1:25.000, tornando sua execução mais rápida e mais barata, sobretudo se houver disponibilidade de dados de sensores remotos de alta resolução espacial a custos mais baixos e facilidade de obtenção dos atributos do terreno nos sistemas de informação geográfica (SIG).

Termos de indexação: pedologia, mapeamento de solos, inteligência artificial, sensores remotos, geomorfometria.

⁽¹⁾ Recebido para publicação em setembro de 2010 e aprovado em março de 2011.

⁽²⁾ Pesquisador A da Embrapa Solos. Rua Jardim Botânico 1024, Jardim Botânico, CEP 22460-000 Rio Janeiro (RJ). E-mails: chagas.rj@gmail.com; waldircj@cnpes.embrapa.br; silvio@cnpes.embrapa.br

SUMMARY: INTEGRATION OF QUICKBIRD DATA AND TERRAIN ATTRIBUTES FOR DIGITAL SOIL MAPPING BY ARTIFICIAL NEURAL NETWORKS

This study evaluated different environmental variables in the digital soil mapping of an area in the northern region of Minas Gerais State, using artificial neural networks. The environmental variables terrain attributes (slope and compound topographic index), the quickbird bands 1, 2 and 3, and lithology were evaluated. The importance of each of the variables in the classification was tested. The “Java Neural Network Simulator” was used with the backpropagation learning algorithm. For each dataset a neural network was created to predict the soil mapping units, and the map produced by the nets was compared with the conventional, to show the general accuracy of each one. The best classification was achieved when all variables were used, with an accuracy of 67.4 % compared to the the conventional soil map. Of the variables, slope was most significant, because when excluded from the dataset, the classification was worst (accuracy 33.7 %). This result showed that the approach can contribute to overcome some problems of soil mapping in Brazil, especially at scales larger than 1:25,000, with faster and cheaper execution, mainly if remote sensing data with high spatial resolution and an affordable price are available, and good digital elevation models to generate the terrain attributes in the geographical information systems.

Index terms: pedology, digital soil mapping; artificial intelligence, remote sensing, geomorphometry.

INTRODUÇÃO

A demanda por mapas de solos de alta resolução para subsidiar o planejamento e a modelagem ambiental tem crescido mundialmente. Entretanto, no Brasil, os recursos financeiros destinados à realização de levantamentos de solos têm decrescido grandemente, embora apenas uma porção muito reduzida do território brasileiro apresente levantamentos em escalas maiores que 1:25.000.

Tradicionalmente utilizadas nos levantamentos de solos convencionais, as fotografias aéreas têm se mostrado ao longo de décadas uma valiosa fonte de dados, porém sua obtenção tem custo elevado, o que torna inviável seu emprego em pequenas áreas. Por sua vez, dados de sensores remotos de alta resolução espacial estão se tornando cada vez mais populares devido ao custo acessível, além de algumas outras vantagens, notadamente o elevado nível de detalhe espacial e as propriedades multiespectrais. Particularmente, os dados do satélite Quickbird, com resolução espacial de 0,6–2,4 m, podem ser uma alternativa às fotografias aéreas convencionais e servir como base para o mapeamento em escalas maiores que 1:25.000 (Manning, 2007).

Novas estratégias e métodos estão sendo desenvolvidos para obtenção de informações sobre solos, utilizando a modelagem para reduzir os custos dos mapeamentos de solos convencionais. Nesse contexto, métodos ou técnicas de mineração de dados – como redes neurais artificiais (RNAs), árvores de decisão e árvores de classificação, entre outros – podem fornecer

soluções para auxiliar na extração de informações a partir de um conjunto de dados existentes (Behrens et al., 2005).

RNA é uma técnica de inteligência artificial que tenta construir um modelo matemático que supostamente trabalha de maneira análoga à do cérebro humano. Em uma RNA existem neurônios para a recepção dos dados, outros para o seu transporte e armazenamento, e um terceiro grupo para a divulgação externa da informação. As RNAs visam adquirir o conhecimento a partir de exemplos, que podem ser usados para uma predição subsequente. Como é necessária a informação *a priori* para o aprendizado da RNA, esse método é conhecido como aprendizado supervisionado (Key et al., 1989; Zell, 1996). De modo geral, o desenvolvimento de um modelo de RNA requer a seleção de dados de treinamento, a escolha de uma arquitetura adequada e um exaustivo e cuidadoso processo de treinamento e validação (Boruvka & Penizek, 2007).

As RNAs têm sido aplicadas na ciência do solo, principalmente para estimativa de atributos do solo. Sua utilização tem sido focada em estudos de propriedades hidráulicas (Minasny et al., 2004), propriedades químicas (Patel et al., 2002) e erosão (Ermini et al., 2005). A aplicação no mapeamento de classes de solos é mais rara e foi relatada em poucos estudos. Zhu (2000) utilizou um conjunto de variáveis ambientais como fonte de dados auxiliares em uma RNA, e a informação espacial derivada do uso dessa abordagem revelou mais detalhes e mostrou qualidade superior à dos derivados do mapa de solos convencional.

RNAs foram utilizadas por Behrens et al. (2005) para predição de unidades de solos, tendo como dados de entrada (variáveis ambientais) 69 diferentes atributos do terreno, 53 unidades geológicas e três tipos de uso da terra extraídos de mapas e bancos de dados existentes. No geral, a abordagem utilizada mostrou-se bastante satisfatória, economizando tempo, reduzindo custos financeiros e produzindo resultados confiáveis.

Boruvka & Penizek (2007) utilizaram RNAs para o mapeamento de unidades de solos. As variáveis ambientais utilizadas foram dados de pH, teor de argila e gradiente textural de levantamentos de solos preexistentes e dados de elevação, aspecto e declividade. Os resultados mostraram a existência de diferenças no sucesso da predição entre as classes de solos avaliadas, estando essa diferença relacionada com a heterogeneidade de cada classe. De modo geral, os atributos do terreno não melhoraram significativamente a predição das classes; em muitos casos, esta foi até pior quando foram combinados os atributos com os dados de solos, em comparação com o uso apenas dos dados de solos.

Chagas et al. (2010) utilizaram nove variáveis ambientais (atributos do terreno, dados do Landsat 7 e um mapa geológico) em uma abordagem por RNAs no mapeamento digital de solos. Essas variáveis foram combinadas e avaliadas quanto à capacidade de discriminação das unidades de solo. Os testes

estatísticos realizados mostraram que as RNAs produziram resultados bastante satisfatórios, demonstrando que essa abordagem pode contribuir para tornar o mapeamento de solos mais rápido e de menor custo.

O objetivo deste trabalho foi avaliar a eficiência dos dados do sensor Quickbird e de atributos do terreno no mapeamento da distribuição dos solos utilizando redes neurais artificiais, tendo em vista a possibilidade de sua utilização como ferramenta auxiliar na elaboração de levantamentos de solos de grande escala (> 1:25.000).

MATERIAL E MÉTODOS

O estudo foi desenvolvido em uma área de aproximadamente 1.440 ha na bacia hidrográfica do rio Itacambiruçu, afluente do rio Jequitinhonha, Estado de Minas Gerais, entre as coordenadas UTM aproximadas de 727.000 e 732.000 mE e 8.161.500 e 8.167.000 mN, fuso 23, Datum Córrego Alegre (Figura 1). A área apresenta clima predominantemente tropical de natureza continental do tipo Aw, de acordo com a classificação de Köppen, isto é, clima tropical, com inverno seco e verão chuvoso, cujo mês mais frio apresenta temperatura média superior a 18 °C, e o mês mais seco, precipitação inferior a 60 mm.

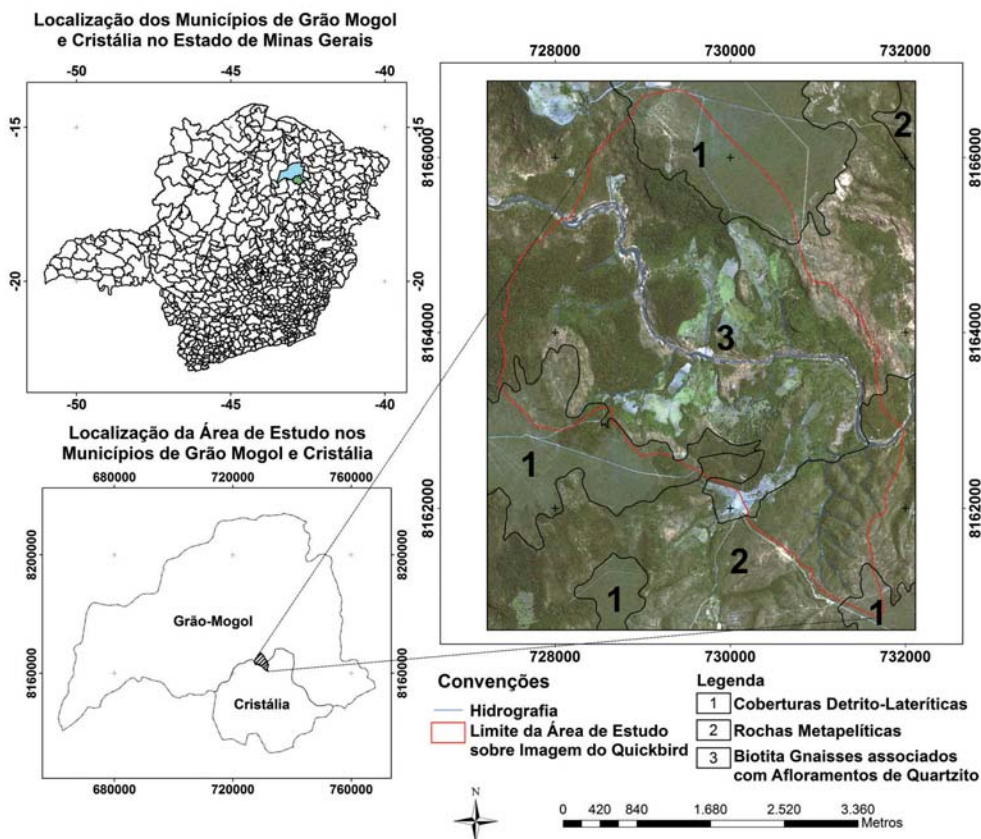


Figura 1. Localização da área de estudo.

Para possibilitar a avaliação dos resultados obtidos com o mapeamento digital de solos, foi elaborado um mapa de solos da área, na escala de 1:20.000, utilizando o método convencional de levantamentos de solos (Embrapa, 1995). Na ausência de fotografias aéreas em escala compatível com os objetivos do estudo ($\geq 1:20.000$), os limites das unidades de mapeamento foram delineados sobre um mapa planialtimétrico, na escala de 1:5.000, com curvas de nível equidistantes de 5 m, e posteriormente digitalizados no ArcGIS 9.3. Os trabalhos de campo foram realizados em fevereiro de 2007 e complementados em agosto de 2008, quando foram descritos e amostrados, conforme Santos et al. (2005), 14 perfis de solo e 10 amostras extras, além de 63 observações de campo.

Nas amostras de solos coletadas, foram efetuadas as seguintes análises, conforme Embrapa (1997): composição granulométrica da terra fina; argila dispersa em água; pH em H₂O e pH em KCl; bases trocáveis (Ca²⁺, Mg²⁺, K⁺, Na⁺); soma de bases (valor S); acidez; capacidade de troca de cátions (valor T); percentagem de saturação por bases (valor V); percentagem de saturação por Al e P disponível; C orgânico; N total; e óxidos do ataque sulfúrico (SiO₂, Fe₂O₃, Al₂O₃, TiO₂, MnO e P₂O₅). Finalmente, os solos foram classificados de acordo com o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006), conforme apresentado na figura 2.

No mapeamento digital de solos podem ser utilizadas diferentes variáveis ambientais como fonte auxiliar de dados (Quadro 1). Essas variáveis representam os fatores que são reconhecidos por governarem a for-

mação do solo (material de origem, relevo, vegetação e clima), bem como a informação do solo que pode ser recuperada a partir de imagens de sensores remotos ou mapas de solos de pequena escala de uma área ou região de interesse (McBratney et al., 2003; Lagacherie, 2008).

Assim, foram utilizadas seis variáveis ambientais, representadas por atributos do terreno, dados de um sensor remoto de alta resolução e um mapa litológico. Um modelo digital de elevação (MDE) com 5 m de resolução espacial foi gerado, a partir de curvas de nível equidistantes de 5 m e hidrografia, na escala de 1:5.000, obtidas exclusivamente para este estudo, utilizando-se a opção "Topo to Raster" do programa ArcGIS 9.3. Por sua vez, esse MDE foi usado para derivar os atributos do terreno declividade (cm m⁻¹) e índice topográfico combinado (CTI, sigla em inglês), que buscam caracterizar as condições topográficas da área. O índice topográfico combinado (CTI) ou também denominado índice de umidade é um atributo topográfico secundário, sendo definido como uma função da declividade (β) e da área de contribuição por unidade de largura ortogonal à direção do fluxo (A_s), conforme a equação 1:

$$CTI = \ln \left(\frac{A_s}{\tan \beta} \right) \quad (1)$$

Uma imagem do sensor Quickbird, com resolução espacial de 2,4 m no modo multiespectral, de 3/04/2003 foi usada como fonte de dados auxiliares, conforme abordagem empregada por Siart et al. (2009). Foram

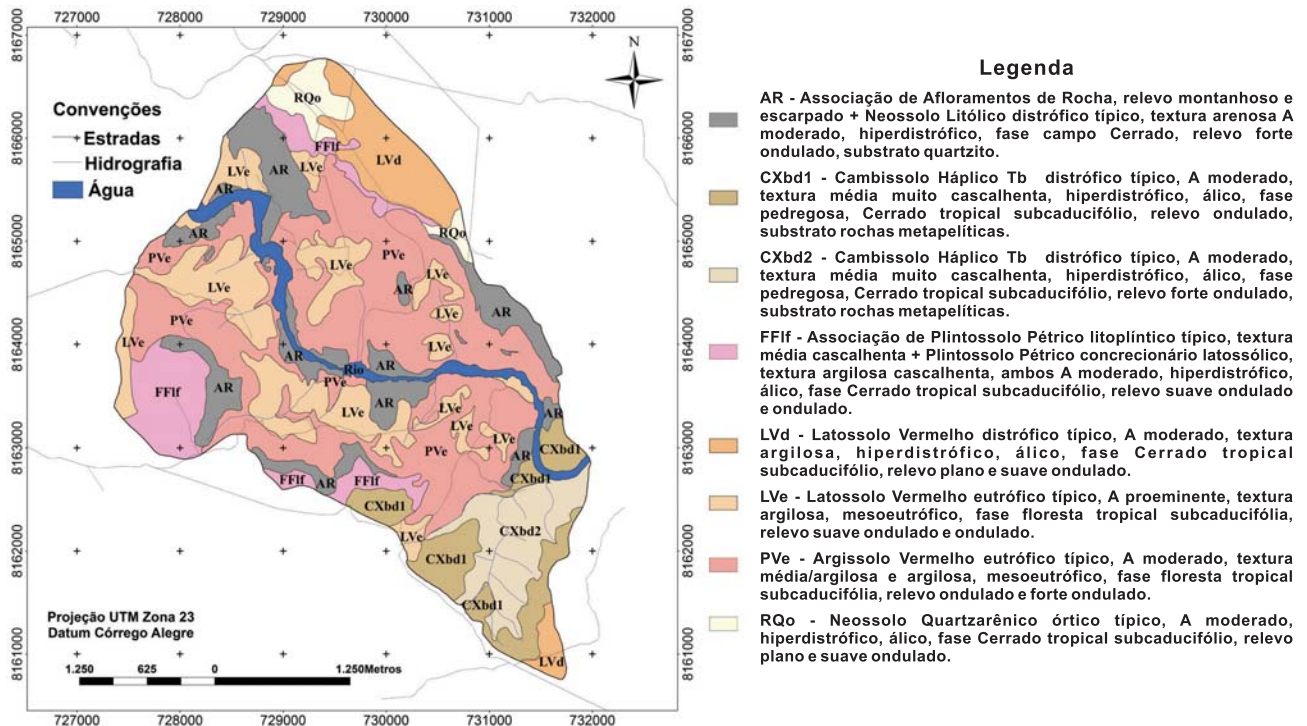


Figura 2. Mapa de solos convencional da área estudada.

Quadro 1. Variáveis ambientais normalmente consideradas em um mapeamento digital de solos e sua significância

Fator de formação	Variável ambiental	Significância	Exemplo de aplicação
Clima (cl)	Precipitação, Temperatura e Radiação Solar	Natureza e taxa de intemperismo e atividade biológica.	Ryan et al. (2000).
Organismos (o)	Bandas de sensores remotos multiespectrais de alta resolução; índice NDVI	Tipo de vegetação, regime de umidade do solo e ciclagem de nutrientes.	Sullivan et al. (2005).
Relevo (r)	Elevação (m)	Variações climáticas, vegetação, energia potencial.	Siart et al. (2009).
	Declividade (cm m ⁻¹)	Taxa de intemperismo/erosão, velocidade e volume de escoamento superficial, teor de água no solo.	Behrens et al. (2005).
	Índice topográfico combinado (CTI)	Zonas de concentração de água na paisagem, regime de umidade do solo.	Chaplot et al. (2000).
Material de Origem (p)	Unidades litológicas	Produtos do intemperismo, incluindo mineralogia e tamanho de partículas.	Thomas et al. (1999).

Adaptado de McKenzie & Ryan (1999).

utilizadas as bandas 1 (0,45–0,52 µm), 2 (0,52–0,60 µm) e 3 (0,63–0,69 µm). Inicialmente, a imagem foi corrigida geometricamente para o sistema de coordenadas UTM (Universal Transversa de Mercator), Datum Córrego Alegre, zona 23S. Em seguida, a imagem foi reamostrada para a mesma resolução espacial das variáveis do terreno (5 m), utilizando-se como método de interpolação o vizinho mais próximo, que usa o valor numérico do pixel de entrada mais próximo para assinalar ao pixel da imagem corrigida. Esse procedimento foi realizado no software ERDAS Imagine 8.5.

Um mapa geológico na escala de 1:100.000 (Oliveira et al., 1997) foi utilizado para dar informações sobre o material de origem dos solos, o qual, com o conhecimento adquirido durante os trabalhos de campo, foi adaptado aos objetivos do estudo (McKenzie & Ryan, 1999). As principais litologias identificadas foram: (a) coberturas detrítico-lateríticas do Terciário, presentes nas partes mais elevadas; (b) rochas metapelíticas pobres, que ocorrem na porção sudeste da área; e (c) biotita gnaisse associada com afloramentos de quartzitos, encontrados na parte central (Figura 1).

As variáveis ambientais foram combinadas no software ERDAS Imagine 8.5. Pelo fato de serem de fontes distintas, buscou-se identificar a importância de cada uma pelo uso dos conjuntos apresentados a seguir: 1 - declividade, CTI, banda 1, banda 2, banda 3 e litologia; 2 - sem a banda 1; 3 - sem a banda 2; 4 - sem a banda 3; 5 - sem o índice CTI; 6 - sem a declividade; e 7 - sem a litologia.

O fluxograma que descreve a abordagem por redes neurais utilizada é mostrado na figura 3.

A utilização de uma RNA inicia-se com o processo de treinamento (Tso & Mather, 2009). Nesse estágio, a rede é treinada para aprender as condições nas quais uma determinada feição ou classe de solo ocorre. Cada unidade de entrada da RNA (neurônio) representa uma variável ambiental preditora (atributos do terreno, dados de sensores remotos, unidades geológicas, etc.), e as unidades de saída representam as classes que se deseja mapear.

Inicialmente, todos os valores referentes às variáveis ambientais foram reescalados para o intervalo entre 0 e 1, utilizando-se o ArcGIS 9.3, no caso dos atributos do terreno, e o ERDAS Imagine 8.5, no caso das bandas do Quickbird. Esse procedimento foi adotado para facilitar o processo de treinamento e

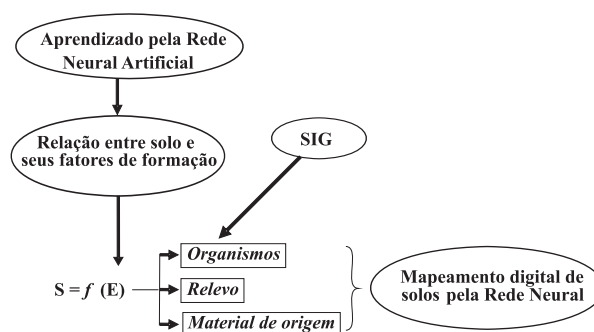


Figura 3. Fluxograma do modelo utilizado na abordagem por redes neurais artificiais. Adaptado de Zhu (2000).

evitar a saturação das RNAs, pois valores muito grandes poderiam impedir a solução do problema (convergência da rede). Além disso, esse procedimento pode prevenir que grandes variações de uma variável pouco importante inibam pequenas variações em outras variáveis.

O simulador de redes neurais utilizado foi o “Java Neural Network Simulator”, que é baseado no “Stuttgart Neural Network Simulator 4.2 Kernel” (Zell et al., 1996). Para cada conjunto de variáveis ambientais, foram coletadas, independentemente, amostras para o treinamento e amostras para a validação das RNAs, que também são independentes entre os conjuntos. Estas amostras (amostras estratificadas) foram coletadas com o auxílio do software ERDAS Imagine 8.5, de maneira a representar o máximo possível às características de cada unidade de mapeamento, com relação às variáveis utilizadas (declividade, índice CTI e bandas do Quickbird). Em seguida, os arquivos foram convertidos para o formato de trabalho do “Java Neural Network Simulator”. Assim, para cada conjunto de variáveis ambientais testado, foi criado um arquivo de treinamento e outro de validação.

Embora Congalton & Green (1999) tenham desenvolvido procedimentos estatísticos para definir o número mínimo de amostras para classificadores, optou-se por seguir as recomendações de Zhu (2000), que, em seu estudo com RNAs, indica que o número de amostras de treinamento deve estar relacionado ao número de unidades de mapeamento ou classes de solos existentes em uma determinada área, pois, quanto mais unidades, mais complexo será o mapeamento da relação entrada (variáveis)/saída (unidades de mapeamento) e, por sua vez, mais amostras serão necessárias para treinar as redes. Assim, o tamanho da amostra de treinamento foi de aproximadamente 30 vezes o número de unidades de mapeamento (300 pixels por unidade de mapeamento), enquanto o conjunto de validação foi 50 % do tamanho do conjunto de treinamento (150 pixels).

Diferentes arquiteturas foram testadas, tendo-se variado o número de neurônios na camada de entrada, correspondente ao número de variáveis ambientais utilizadas (diferentes conjuntos), bem como o número de neurônios na camada interna, e todas contendo o mesmo número de neurônios na camada de saída, correspondente ao número de unidades de mapeamento (8). O número de neurônios da camada interna foi determinado por meio de tentativa e erro, conforme sugerido por Hirose et al. (1991). O critério adotado para adicionar neurônios à camada interna foi baseado no comportamento do erro quadrático médio (EQM) durante essa fase, que mede a diferença entre os valores estimados e os valores desejados para o treinamento, conforme a equação 2.

$$EQM = \frac{1}{n} \sum (e - d)^2 \quad (2)$$

em que e representa os valores estimados para cada pixel; e d , os valores desejados. Assim, o treinamento deve ser interrompido quando esse erro, para o conjunto de treinamento testado, tende ao menor possível e não mais oscila com novos ciclos de treinamento.

No treinamento das RNAs, foi utilizado o algoritmo de aprendizado “backpropagation”, com alocação aleatória dos pesos interneurônios entre -0,5 e 0,5 e uma taxa de aprendizado de 0,2, considerando 10 mil ciclos de aprendizagem.

A avaliação dos resultados foi realizada com medidas estatísticas, como o índice Kappa e a exatidão global da classificação, derivados de uma matriz de confusão (Congalton & Green, 1999). Uma matriz de significância foi gerada com os resultados dos testes estatísticos, tendo-se utilizado os valores de Kappa e de sua variância entre as classificações. O teste estatístico Z verifica, inicialmente, se a classificação difere de uma classificação casual e, em segunda análise, se existe diferença significativa entre os valores de Kappa resultantes da avaliação dos diferentes conjuntos, aos pares. Ao término do processo de treinamento e validação das RNAs, foi selecionada, para o mapeamento digital das classes de solos da área, aquela arquitetura de rede que obteve o melhor resultado para o índice Kappa, para cada um dos conjuntos de variáveis ambientais utilizados.

Por fim, os mapas de solos digitais produzidos por redes neurais para cada combinação de variáveis ambientais foram comparados com o mapa de solos convencional (Figura 2), onde foi determinada a porcentagem de concordância de cada mapa digital com o mapa de solos convencional.

RESULTADOS E DISCUSSÃO

Para todos os conjuntos de variáveis ambientais, não foram verificadas diferenças significativas nos valores do EQM entre as redes a partir de 10 neurônios em uma única camada interna. Assim, a fase de treinamento foi encerrada e, para cada conjunto, foi escolhida a topologia de rede com 10 neurônios na camada interna, cujos resultados para o EQM variaram de 0,16 (sem a declividade) a 0,02 (com todas as variáveis). Segundo Faussett (1994), teoricamente, uma topologia contendo apenas uma camada interna tem sido suficiente para extrair conhecimento relevante a partir de um conjunto de dados de treinamento. Além disso, de acordo com Foody & Arora (1997), em geral, redes maiores e mais complexas são mais eficientes para caracterizar corretamente um conjunto de treinamento, porém podem ser menos eficientes do que redes mais simples para generalizar.

A influência do número e categoria das variáveis ambientais, com relação ao desempenho das redes, utilizando-se as amostras de validação, é apresentada no quadro 2.

Quadro 2. Matriz de significância de Kappa para os conjuntos de variáveis ambientais testados, utilizando-se as amostras de validação

Conjunto	1 - Todas	2 - Sem a Banda 1	3 - Sem a Banda 2	4 - Sem a Banda 3	5 - Sem o Índice CTI	6 - Sem a declividade	7 - Sem a Litologia
Exatidão global	89,3	85,2	84,5	84,2	85,1	62,0	52,3
Kappa	0,881	0,836	0,828	0,824	0,835	0,578	0,470
Variância	0,000078	0,000104	0,000107	0,000109	0,000104	0,000193	0,000203
1 - Todas	99,75						
2 - Sem a banda 1	3,34*	81,98					
3 - Sem a banda 2	3,90*	0,55	80,05				
4 - Sem a banda 3	4,17*	0,82	0,27	78,93			
5 - Sem o índice CTI	3,41*	0,07	0,48	0,75	81,88		
6 - Sem a declividade	18,41*	14,97*	14,43*	14,16*	14,91*	41,61	
7 - Sem a litologia	24,52*	20,89*	20,33*	20,04*	20,83*	5,43*	32,99

existe diferença significativa a 5 %.

A matriz de significância de Kappa mostra que a combinação de todas as variáveis foi a que apresentou o melhor desempenho, com valor de 0,881 para o índice Kappa, considerado por Landis & Koch (1977) como muito bom a excelente. Semelhantemente a outros estudos, o uso combinado de atributos do terreno, dados de sensores remotos e informações sobre litologia como preditores obteve os melhores resultados (Behrens et al., 2005; Chagas et al., 2010).

Por outro lado, os piores desempenhos foram obtidos quando a declividade (Kappa de 0,578) e a litologia (Kappa de 0,470) foram removidas da análise. Esses desempenhos são considerados moderados e diferem estatisticamente, a 5 % do intervalo de confiança, do conjunto 1. Os conjuntos 2 (sem a banda 1), 3 (sem a banda 2), 4 (sem a banda 3) e 5 (sem o índice CTI) também tiveram desempenho de muito bom a excelente

para o índice Kappa, e não diferem estatisticamente do conjunto que emprega todas as variáveis, assim como não diferem entre si (Quadro 2).

As diferenças das variáveis ambientais entre as unidades de mapeamento, considerando as amostras utilizadas no treinamento das redes, são apresentadas na figura 4.

A figura 4 mostra que as diferenças entre as unidades de mapeamento são mais pronunciadas para os atributos do terreno do que as verificadas para as bandas do Quickbird. Dos atributos do terreno, a declividade é o que mostra comportamento mais diferenciado entre as unidades de mapeamento e, certamente, contribuiu com maior peso para o desempenho da classificação, já que sua retirada da análise interferiu negativamente na exatidão global (Quadro 2).

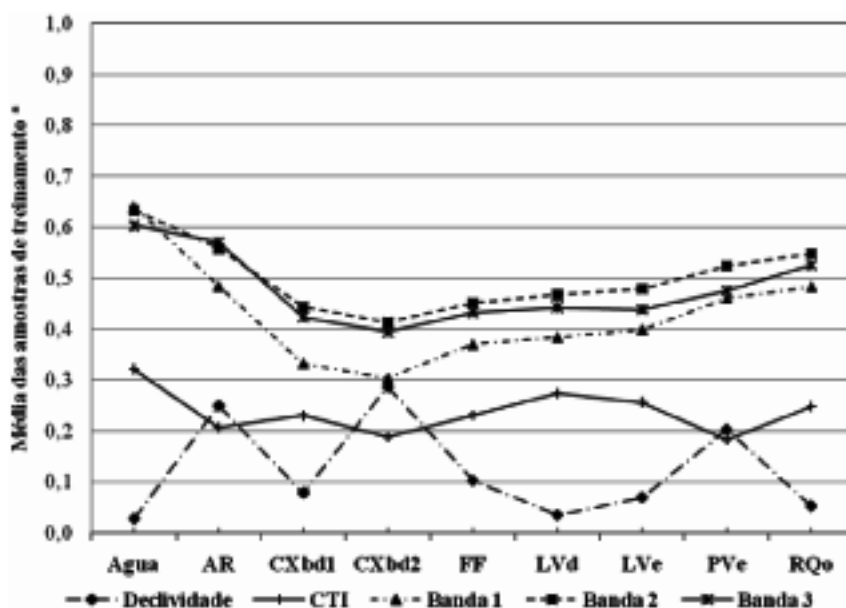


Figura 4. Distribuição das médias das variáveis ambientais entre as unidades de mapeamento. *: valores reescalados entre 0 e 1 para permitir a apresentação em conjunto.

O índice CTI apresenta diferenciação menos pronunciada do que a declividade, entre as unidades de mapeamento, com valores que variam pouco, com exceção da água; dessa forma, contribui com menor peso para a separabilidade das unidades, o que foi comprovado pelo índice Kappa ligeiramente inferior (estatisticamente diferente) ao obtido com a utilização de todas as variáveis, quando essa variável foi retirada da análise. Valores elevados do índice CTI estão relacionados normalmente a áreas planas de baixada ou áreas côncavas, que favorecem o acúmulo de água no solo (Penizek & Boruvka, 2008). A forte influência dos atributos do terreno sobre a distribuição espacial dos solos é um princípio bem conhecido (Park & Vlek, 2002), o que também foi confirmado neste estudo pela diminuição dos valores de Kappa quando estes foram retirados da análise, principalmente a declividade.

As bandas do Quickbird, especialmente a 2 e a 3, mostraram comportamento muito parecido entre si e entre as unidades de mapeamento (Figura 4), contribuindo com menor peso para o desempenho do treinamento, quando foram utilizadas as amostras de validação, o que é comprovado pelos resultados do índice Kappa, quando estas bandas foram suprimidas da análise (Quadro 2), que foram ligeiramente inferiores aos obtidos com a utilização de todas as variáveis. Embora a banda 1 tenha apresentado comportamento ligeiramente mais diferenciado, sobretudo para as unidades Água, AR, CXbd1 e CXbd2, também contribuiu pouco para melhorar o desempenho do treinamento (Quadro 2). Segundo Dobos et al. (2001), dados de sensores remotos são fortemente influenciados pela variabilidade do terreno e ainda não conseguem identificar com exatidão toda a variabilidade do solo que ocorre na paisagem; por isso, devem ser complementados com informações do terreno para compensar as distorções que surgem das variações topográficas da paisagem, além de fornecerem dados adicionais para modelagem solo-paisagem.

No quadro 3 e na figura 5 são apresentados os resultados da análise de concordância entre os mapas digitais produzidos pelos diferentes conjuntos e o mapa de solos convencional. Embora os resultados obtidos para o índice Kappa, quando se utilizaram as amostras

de validação, tenham variado de moderado a excelente, de acordo com Landis & Koch (1977), os mapas de solos produzidos pelos diferentes conjuntos de variáveis mostraram-se distintos.

A maior concordância com o mapa de solos convencional foi obtida quando se utilizaram todas as variáveis ambientais (atributos do terreno, as bandas do Quickbird e a litologia), com 67,4 % de concordância, corroborando os resultados obtidos para o índice Kappa (Quadro 2) durante a fase de treinamento das redes. Já o pior desempenho foi obtido quando a declividade foi removida da análise (conjunto 6), resultando numa concordância de apenas 33,7 %, indicando que a declividade, entre as variáveis utilizadas, é a que mais contribui para melhorar o desempenho da classificação dos dados. As variáveis que aparentemente menos contribuíram foram as bandas 1 e 2, com 16,6 e 14,9 % de contribuição, respectivamente (Quadro 3). A retirada da banda 3, do índice CTI e da litologia das análises produziu resultados semelhantes na comparação com o mapa de solos convencional, com valores próximos a 45 %, indicando contribuição similar para o desempenho da classificação.

Os resultados obtidos quando as variáveis banda 1, banda 2, banda 3 e índice CTI (Quadro 3) foram retiradas da análise sugerem que durante a fase de treinamento ocorreu *overfitting*, ou seja, as redes se especializaram nos padrões de treinamento e tiveram sua capacidade de generalização ou de classificação dos dados comprometida (Sarle, 1995), já que os resultados do índice Kappa foram bastante satisfatórios (> 0,80) e muito próximos dos obtidos quando todas as variáveis foram utilizadas (Quadro 2).

As diferentes unidades de mapeamento tiveram concordâncias com o mapa de solos convencional, distintas entre os conjuntos de variáveis ambientais testados (Quadro 3). As unidades água, CXbd1, CXbd2 e FFIf tiveram maior concordância quando todas as variáveis foram utilizadas; esses resultados foram consideravelmente superiores aos obtidos pelos demais conjuntos.

A unidade AR apresentou a maior concordância quando a banda 1 foi retirada da análise (49,0 %),

Quadro 3. Resultados da análise de concordância entre os mapas digitais e o mapa convencional

Conjunto	Acerto	Contribuição da variável	Acerto nas unidades de mapeamento								
			Água	AR	CXbd1	CXbd2	FFIf	LVd	LVe	PVe	RQo
%											
1 - Todas	67,4	-	85,1	45,6	82,3	86,7	63,4	80,8	94,8	52,5	50,8
2 - Sem a banda 1	50,8	16,6	0,3	49,0	10,2	36,2	22,5	80,2	92,4	50,8	21,0
3 - Sem a banda 2	52,5	14,9	0,4	41,0	2,0	41,4	50,8	89,0	91,1	53,3	3,6
4 - Sem a banda 3	44,3	23,1	62,8	35,1	0,3	37,5	4,7	30,6	92,6	47,2	76,7
5 - Sem o índice CTI	44,4	23,0	10,9	44,3	3,2	2,5	30,8	11,6	94,1	53,9	0,0
6 - Sem a declividade	33,7	33,7	10,9	27,0	1,3	4,6	0,7	0,0	76,5	47,6	0,0
7 - Sem a litologia	42,8	24,6	24,6	9,1	0,3	44,8	45,6	61,5	82,9	45,3	2,2

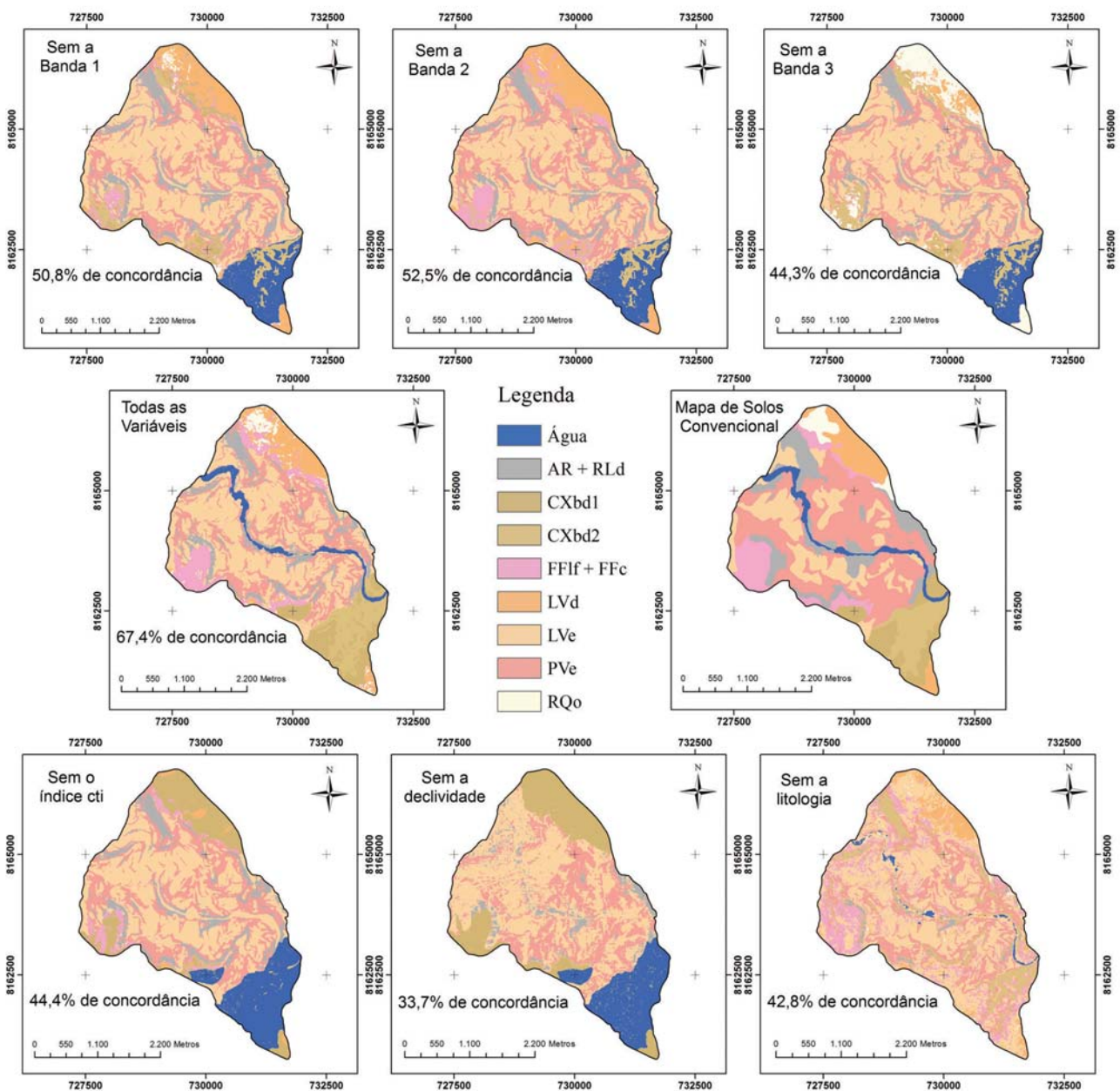


Figura 5. Mapas digitais de solos produzidos pelos diferentes conjuntos de variáveis ambientais.

seguida da análise que utilizou todas as variáveis, com uma concordância ligeiramente inferior (45,6 %). Os demais conjuntos apresentaram comportamento similar, com exceção do conjunto 7, em que a remoção da litologia produziu concordância de apenas 9,1 %. Nesse caso específico, esperava-se contribuição mais significativa das bandas do Quickbird, visto que visualmente a identificação desta unidade é bastante nítida, o que acabou não acontecendo.

A maior concordância com o mapa convencional, para a unidade LVe, foi obtida quando se retirou a banda 2, com 89 %, seguida pelo conjunto que emprega todas as variáveis (80,8 %) e pelo conjunto que não utiliza a banda 1 (80,2 %). A remoção da banda 3

(conjunto 4), do índice CTI (conjunto 5) e da declividade (conjunto 6) interferiu negativamente nos resultados para esta unidade, com concordâncias inferiores a 31 %. No caso da remoção da declividade, essa interferência foi ainda maior, já que não houve concordância (0 %) com o mapa convencional.

A unidade de mapeamento LVe obteve concordâncias bastante similares entre os conjuntos, superiores a 80 % para quase todos eles, exceto quando se retirou a declividade (76,5 %). Esses resultados indicam que esta unidade é pouco influenciada pelas variáveis de um modo isolado. A maior concordância foi obtida quando todas as variáveis foram utilizadas (94,8 %). A unidade PVe teve comportamento, com

relação aos diferentes conjuntos, semelhante ao verificado para a unidade LVe, sendo a concordância com o mapa convencional, no entanto, inferior (Quadro 3), em função do confundimento com a unidade AR. A maior concordância foi obtida sem a utilização do índice CTI (conjunto 5), com 53,9 %, ligeiramente superior ao conjunto 3 (53,3 %).

Já a unidade de mapeamento RQo obteve o melhor resultado quando a banda 3 não foi utilizada, com 76,7 % de concordância com o mapa convencional, seguida pelo conjunto 1 (50,8 %). Os demais conjuntos obtiveram concordâncias muito baixas (< 21 %), assim como verificado para a unidade LVd.

Considerando apenas a classificação pelo conjunto 1 (todas as variáveis), que obteve os melhores resultados, verifica-se que as diferenças mais marcantes observadas entre o mapa digital e o mapa convencional ocorreram para as unidades AR (45,6 %) e RQo (50,8 %). No quadro 4 é apresentado o confundimento ocorrido entre as unidades de mapeamento para a classificação utilizando-se todas as variáveis. A unidade AR apresentou maior confusão com a unidade PVe, com 24,0 % dos pixels que deveriam ser classificados como pertencentes à unidade AR sendo atribuídos a esta classe. Por outro lado, 10,8 % dos pixels que deveriam ser atribuídos à unidade PVe foram atribuídos à unidade AR. Uma possível explicação para isso é a similaridade nas características dos atributos do terreno e litologia entre essas unidades (Figura 4), que interferiu negativamente na classificação, diminuindo consequentemente a percentagem de concordância para ambas. A influência negativa da similaridade entre características de variáveis ambientais sobre a classificação de dados foi relatada por Boruvka & Penizek (2007).

A unidade de mapeamento RQo apresentou alguma confusão com as unidades de mapeamento LVd (22,3 %) e FFIf (15,1 %). Nesses casos, a confusão deve-se basicamente à grande similaridade na declividade e no índice CTI entre elas (Figura 4). Essa confusão ocorre principalmente nos limites entre essas unidades, em razão da natureza contínua dos solos.

Neste caso, embora as bandas do Quickbird apresentem valores distintos entre essas três unidades (Figura 4), elas não foram suficientes para melhorar o desempenho da classificação dessas unidades de mapeamento, evidenciando deficiência das variáveis utilizadas e, ou, do classificador (redes neurais).

As unidades CXbd1 e CXbd2 apresentaram concordância com o mapa convencional muito satisfatória (82,3 e 86,7 %, respectivamente) e tiveram pequeno confundimento entre si, com 11,6 % dos pixels que deveriam ser classificados como pertencentes à unidade CXbd1 sendo assinalados para a classe CXbd2. Por sua vez, 14,1 % dos pixels que deveriam ser classificados como pertencentes à unidade CXbd2 foram erroneamente classificados como pertencentes à classe CXbd1. Essa confusão deve-se ao fato de que essas unidades de mapeamento apresentam características muito similares para a maioria das variáveis discriminantes utilizadas, com exceção da declividade (Figura 4).

Diferenças marcantes das variáveis ambientais (Figura 4) resultaram em um baixo confundimento da unidade de mapeamento FFIf com as demais, sendo de 8,8 % com a unidade de mapeamento LVd; de 6,7 % com a unidade AR; de 6,2 % com a unidade PVe e de 6,4 % com a unidade RQo. Do mesmo modo, a unidade LVd apresentou ligeira confusão com as unidades FFIf (10,2 %) e RQo (8,3 %). Esses confundimentos na classificação são observados principalmente nos limites entre essas unidades, que ocorrem todas em áreas contíguas na cobertura detrito-lateríticas do Terciário.

Outra confusão – e, com certeza, a mais importante – é a que ocorreu entre as unidades LVe e PVe, com 38,3 % dos pixels que deveriam ser classificados como pertencentes à unidade LVe sendo atribuídos à unidade PVe. Nesse caso, os pixels classificados erroneamente na unidade PVe também estão localizados nos limites entre essas unidades, o que configura certa dificuldade do classificador (redes neurais) em estabelecer a correta classificação para pixels que possuem características ambientais muito próximas, principalmente a declividade. A unidade LVe ocorre em áreas de relevo suave ondulado e

Quadro 4. Confundimento entre as unidades de mapeamento para a classificação pelo conjunto 1

Unidade	Água	AR	CXbd1	CXbd2	FFIf	LVd	LVe	PVe	RQo
	%								
Água	91,7	0,5	0,4	2,2	0,2	0,03	0,4	0,5	1,1
AR	5,0	71,5	0,2	2,4	6,7	0,2	11,5	10,8	0,1
CXbd1	0,7	0,0	82,4	14,1	0,6	0,3	0,4	0,0	1,0
CXbd2	0,1	0,0	11,6	78,4	0,0	0,0	0,1	0,0	0,0
FFIf	0,0	3,3	3,3	0,2	69,5	10,2	2,3	2,3	15,1
LVd	0,0	0,0	2,0	0,0	8,8	80,1	0,0	0,0	22,3
LVe	0,9	0,4	0,0	0,0	1,5	0,8	46,4	3,0	0,7
PVe	1,6	24,0	0,2	2,5	6,2	0,1	38,3	83,0	0,5
RQo	0,0	0,3	0,0	0,0	6,4	8,3	0,6	0,4	59,2
Total de pixels classificados	77.866	197.566	173.917	177.662	171.375	164.989	819.458	504.051	56.066

ondulado, enquanto a unidade PVe ocupa áreas de relevo ondulado e forte ondulado. Não se pode, também, ignorar possíveis erros aleatórios no mapa de solos convencional utilizado para a comparação, oriundos da delimitação manual das unidades de mapeamento, o que certamente influencia os resultados obtidos (Behrens et al., 2005).

As possíveis causas de discordância entre as classificações podem estar relacionadas com: ineficiência no aprendizado/treinamento das RNAs, a partir dos conjuntos de dados testados; deficiência das próprias variáveis ambientais em representar as variações das características das unidades de mapeamento (Behrens et al., 2005); e problemas no estabelecimento de um modelo de correlação ambiental mais realista e adequado às condições da área, que, embora possa ser empregado para predição de solos, pode não utilizar toda a capacidade preditiva dos modelos mentais intuitivos usados no levantamento convencional (McKenzie & Ryan, 1999).

CONCLUSÕES

1. A utilização conjunta dos dados do sensor Quickbird, dos atributos do terreno (declividade e índice CTI) e do mapa geológico em uma abordagem por redes neurais artificiais obteve resultados satisfatórios, tanto no processo de treinamento das redes (exatidão global de 89,3 % e índice Kappa de 0,881) quanto na comparação com o mapa de solos convencional (concordância de 67,4 %), podendo ser utilizada como uma alternativa viável de execução mais rápida e de menor custo no mapeamento de solos de grande escala.

2. A variável ambiental que mais contribuiu para melhorar o desempenho da classificação dos dados foi a declividade, que apresenta comportamento mais distinto entre as unidades de mapeamento da área.

3. As possíveis causas de discordância entre o mapa de solos convencional e o mapa de solos digital podem estar relacionadas com: deficiências no aprendizado dos critérios utilizados no treinamento das RNAs; deficiências das variáveis ambientais utilizadas em representar todas as variações das características ambientais das unidades de mapeamento; e deficiências no estabelecimento de um modelo de relação solo-paisagem mais adequado às condições da área – aspecto que enfatiza a necessidade dos estudos de campo para o correto entendimento dessas relações.

LITERATURA CITADA

- BEHRENS, T.; FOSTER, H.; SCHOLTEN, T.; STEINRUCKEN, U.; SPIES, E.D. & GOLDSCHMITT, M. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.*, 168:21-33, 2005.
- 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.B. & VOLTZ, M., eds. *Digital soil mapping: An introductory perspectives*. Amsterdam, Elsevier, 2007. p.415-424. (Developments in Soil Science, 31)
- CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.R. & CARVALHO JUNIOR, W. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. *Pesq. Agropec. Bras.*, 45:497-507, 2010.
- CHAPLOT, V.; WALTER, C. & CURMI, P. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 97:405-422, 2000.
- CONGALTON, R.G. & GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. New York, Lewis Publishers, 1999. 137p.
- DOBOS, E.; MONTANARELLA, L.; NÈGRE, T. & MICHELI, E. A regional scale soil mapping approach using integrated AVHRR and DEM data. *Intern. J. Appl. Earth Observ. Geoinform.*, 3:30-42, 2001.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. *Procedimentos normativos de levantamentos pedológicos*. Brasília, Embrapa - SPI, 1995. 101p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. *Sistema brasileiro de classificação de solos*. 2.ed. Rio de Janeiro, 2006. 306p.
- ERMINI, L.; CATANI, F. & CASAGLI, N. Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology*, 66:327-343, 2005.
- FAUSSETT, L.V. *Fundamentals of neural networks: Architectures, algorithms and applications*. New Jersey, Prentice Hall, 1994. 461p.
- FOODY, G.M. & ARORA, M.K. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *Intern. J. Remote Sens.*, 18:799-810, 1997.
- HIROSE, Y.; YAMASHITA, K. & HIJIYA, S. Back-propagation algorithm which varies the number of hidden units. *Neural Networks*, 4:61-66, 1991.
- KEY, J.; MASLANIK, J.A. & SCHWEIGER, A.J. Classification of merged AVHRR and SMMR arctic data with neural networks. *Photogram. Eng. Remote Sens.*, 55:1331-1338, 1989.
- LAGACHERIE, P. Digital soil mapping: A state of the art. In: HARTEMINK, A.E.; McBRATNEY, A.B. & MENDONÇA-SANTOS, M.L., eds. *Digital soil mapping with limited data*. Dordrecht, Springer, 2008. p.3-14.
- LANDIS, J.R. & KOCH, G.G. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174, 1977.

- MANNING, J. Remote sensing for terrain analysis of linear infrastructure projects. In: TEEUW, R., ed. Mapping hazardous terrain using remote sensing. Geol. Soc. London, 283:135-142, 2007. (Special Publications)
- McBRATNEY, A.B.; SANTOS, M.L.M. & MINASNY, B. On digital soil mapping. *Geoderma*, 117:3-52, 2003.
- McKENZIE, N.J. & RYAN, P.J. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89:67-94, 1999.
- MINASNY, B.; HOPMANS, J.W.; HARTEK, T.; ECHING, S.O.; TULI, A. & DENTON, M.A. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Sci. Soc. Am. J.*, 68:417-430, 2004.
- OLIVEIRA, M.J.R.; GROSSI-SAD, J.H.; ROMANO, A.W. & LOBATO, L.M. Geologia da Folha Grão Mogol. In: GROSSI-SAD, J.H.; LOBATO, L.M.; PEDROSA-SOARES, A.C. & SOARES-FILHO, B.S., eds. Projeto Espinhaço em CD-ROM (textos, mapas e anexos). Belo Horizonte, COMIG - Companhia Mineradora de Minas Gerais, 1997. p.611-713.
- PARK, S.J. & VLEK, P.L.G. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. *Geoderma*, 109:117-140, 2002.
- PATEL, R.M.; PRASHER, S.O.; GOEL, P.K & BASSI, R. Soil salinity prediction using artificial neural networks. *J. Am. Water Res. Assoc.*, 38:91-100, 2002.
- PENIZEK, V. & BORUVKA, L. The digital terrain model as a tool for improved delineation of alluvial soils. In: HARTEMINK, A.E.; McBRATNEY, A.B. & MENDONÇA-SANTOS, M.L., eds. Digital soil mapping with limited data. Dordrecht, Springer, 2008. p.319-326.
- RYAN, P.J.; McKENZIE, N.J.; O'CONNELL, D.; LOUGHHEAD, A.N.; LEPPERT, P.M.; JACQUIER, D. & ASHTON, L. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. *For. Ecol. Manag.*, 138:139-157, 2000.
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H.C. Manual de descrição e coleta de solo no campo. 5.ed. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, 2005. 100p.
- SARLE, W.S. Stopped training and other remedies for overfitting. In: SYMPOSIUM ON THE INTERFACE COMPUTING SCIENCE AND STATISTICS, 27., 1995, Pittsburgh. Proceedings... Pittsburgh, 1995. p.1-10.
- SIART, C.; BUBENZER, O. & EITEL, B. Combining digital elevation data (SRTM/ASTER), high resolution satellite imagery (Quickbird) and GIS for geomorphological mapping: A multi-component case study on Mediterranean karst in Central Crete. *Geomorphology*, 112:106-121, 2009.
- SULLIVAN, D.G.; SHAW, J.N. & RICKMAN, D. Ikonos imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Sci. Soc. Am. J.*, 69:1789-1798, 2005.
- THOMAS, A.L.; KING, D.; DAMBRINE, E. & COUTURIER, A. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France). *Geoderma*, 90:291-305, 1999.
- TSO, B. & MATHER, P.M. Classification methods for remotely sensed data. 2.ed. Boca Raton, CRC, 2009. 356p.
- ZELL, A. et al. Stuttgart Neural Network Simulator v4.2. Stuttgart, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems/University of Tübingen - Wilhelm-Schickard-Institute for Computer Science - Department of Computer Architecture, 1996. 338p.
- ZHU, A.X. Mapping soil landscape as spatial continua: The neural network approach. *Water Res. Res.*, 36:663-677, 2000.