

ELIANA DE SOUZA

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

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

**VIÇOSA
MINAS GERAIS - BRASIL
2009**

Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV

T

S729c
2009

Souza, Eliana de, 1976-

Classificação supervisionada de solos por redes neurais artificiais na Serra do Cipó-MG / Eliana de Souza. – Viçosa, MG, 2009.

xiv, 96f. : il. (algumas col.) ; 29cm.

Inclui anexo.

Orientador: Elpídio Inácio Fernandes Filho.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 81-86

1. Mapeamento do solo. 2. Mapeamento digital. 3. Redes neurais (Computação). 4. Solos - Classificação. 5. Sistemas de informalão geográfica. 6. Ciência do solo. 7. Solos - Sensoriamento remoto. I. Universidade Federal de Viçosa. II. Título.


CDD 22. ed. 631.4

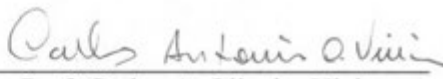
ELIANA DE SOUZA

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

Dissertação apresentada à
Universidade Federal de Viçosa, como
parte das exigências do Programa de Pós-
Graduação em Solos e Nutrição de
Plantas, para obtenção do título de
Magister Scientiae.

APROVADA: 16 de fevereiro de 2009.


Prof. João Carlos Ker
(Co-orientador)


Prof. Carlos A. Oliveira Vieira


César da Silva Chagas


Felipe Nogueira Bello-Simas


Prof. Elpidio-Inácio Fernandes Filho
(Orientador)

*A minha mãe, Geralda pelo carinho,
incentivo e pela disponibilidade sempre
que preciso.*

*A meus irmãos e sobrinhos pela alegria
com que me acolhem,
dedico,*

The beautiful thing about learning is that no one can take it away from you.
B.B. King

AGRADECIMENTOS

À Universidade Federal de Viçosa, por meio do Departamento de Solos e Programa de Pós-Graduação em Solos e Nutrição de Plantas, pela oportunidade de realizar o aperfeiçoamento profissional através do curso de mestrado.

Ao CNPq e à CAPES, pelo apoio com a concessão das bolsas de estudo, e à FAPEMIG, pelo financiamento do projeto.

Ao IBAMA e IEF-MG, em nome dos funcionários do Parque Nacional da Serra do Cipó e APA Morro da Pedreira, pelo apoio nos trabalhos de campo.

Agradeço em especial ao professor Elpídio, pela orientação, pela paciência e amizade durante todo o desenvolvimento da pesquisa.

Ao Professor Carlos Schaefer, pelas orientações desde a Iniciação Científica e pela oportunidade de participar do projeto de estudos que deu início a esta dissertação.

Aos Professores João Ker e Carlos Vieira, pelas orientações e ensinamentos essenciais para o desenvolvimento deste trabalho.

Ao pesquisador da Embrapa Solos, César Chagas, pelas contribuições dadas a este trabalho, pela colaboração nos trabalhos de campo, para coleta de solos, e de laboratório, com as redes neurais.

Ao Felipe Simas, pela amizade e pelas contribuições dadas.

Aos colegas da Pós-Graduação que participaram dos trabalhos de levantamento do meio físico na Serra do Cipó, Roberto Michel, Bruno Nery, Acauã, Elton Valente e Guilherme. Aos demais colegas da Pós-Graduação e Graduação da UFV, pelo apoio.

Aos professores do DPS, pelas contribuições fornecidas durante o decorrer do curso, nas disciplinas cursadas e nos seminários. Aos funcionários do DPS, pelo apoio.

Aos colegas e amigos, pelas experiências, dúvidas e alegrias compartilhadas durante minha vida acadêmica na UFV, desde a graduação. Em especial à Ana Paula de Oliveira, pela amizade, e à Ecíla Villani e ao Esequiel G. da Silva, pelas correções.

À minha família, meus irmãos e, em especial, à minha mãe, pelo carinho, paciência e incentivo.

A todos que de alguma forma contribuiu com este trabalho, **MUITO OBRIGADA...**

BIOGRAFIA

ELIANA DE SOUZA nasceu no dia 04 de agosto de 1976, na cidade de Raul Soares, no estado de Minas Gerais, Brasil. É filha de Geralda Maria de Souza e Sebastião Canuto de Souza (*in memoriam*).

Em 2002, iniciou o curso de Geografia, na Universidade Federal de Viçosa, Minas Gerais, graduando-se em Licenciatura e Bacharelado em março de 2007.

Em março de 2007 iniciou o curso de mestrado no Programa de Pós-Graduação em Solos e Nutrição de Plantas do Departamento de Solos, na UFV, defendendo a dissertação em fevereiro de 2009.

SUMÁRIO

LISTA DE FIGURAS.....	vii
LISTA DE QUADROS.....	viii
LISTA DE ANEXOS.....	x
RESUMO.....	xi
ABSTRACT.....	xiii
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA.....	2
2.1. Mapeamento digital de solos.....	2
2.2. Redes Neurais Artificiais.....	6
2.3. Variáveis discriminantes dos solos.....	9
2.4. Avaliação da exatidão do mapeamento e comparação entre mapas.....	12
3. MATERIAL E MÉTODOS.....	14
3.1. Localização e caracterização da área de estudo.....	14
3.2. Base de Dados.....	17
3.3. Dados de Solos.....	18
3.4. Variáveis discriminantes dos solos.....	21
3.4.1. Modelo Digital de Elevação.....	21
3.4.2. Aspecto.....	23
3.4.3. Declividade.....	24
3.4.4. Amplitude altimétrica.....	26
3.4.5. Curvatura.....	27
3.4.6. Radiação solar.....	29
3.4.7. Índice Topográfico Combinado.....	30
3.4.8. Geologia.....	31
3.4.9. Dados espectrais de imagem de satélite orbitais e índices derivados.....	32
3.4.10. Informação pedológica.....	35
3.5. Amostragem de padrões.....	37
3.6. Treinamento do classificador e definição da arquitetura das redes neurais.....	38

3.7. Validação dos mapas.....	39
3.8. Comparação entre os mapas obtidos por diferentes classificadores.....	45
4. RESULTADOS E DISCUSSÃO.....	47
4.1. Treinamento do classificador.....	52
4.1.1. Número de neurônios.....	53
4.1.2. Ciclos de treinamento na classificação pelas redes neurais.....	55
4.1.3. Avaliação das variáveis discriminantes.....	56
4.1.4. Variáveis discriminantes na classificação pelas redes neurais.....	56
4.1.5. Variáveis discriminantes na classificação pelo MaxVer.....	63
4.2. Comparação entre as classificações feitas pelas Redes Neurais e pelo MaxVer.....	67
4.3. Classificação e pós-processamento dos mapas de solos.....	68
4.4. Validação do mapa de solos com pontos de referência terrestre.....	71
4.4.1. Avaliação das unidades de solos classificadas pelo MaxVer e pelas Redes Neurais.....	72
4.4.2. Validação do mapa gerado pelo método convencional.....	76
5. CONCLUSÕES.....	80
REFERÊNCIAS BIBLIOGRÁFICAS.....	81
ANEXOS.....	87

LISTA DE FIGURAS

Figura 1 – Arquitetura de rede neural. Modificada de Kavzoglu e Mather (2003).....	7
Figura 2 – Fluxograma das etapas de execução do trabalho.	14
Figura 3 – Mapa de localização da área de estudo.....	15
Figura 4 – Mapa de localização dos perfis de solos utilizados.	19
Figura 5 – Fluxograma das variáveis discriminantes utilizadas.....	21
Figura 6 – Mapa de elevação do terreno.....	22
Figura 7 – Mapa de aspecto.....	24
Figura 8 – Mapa de declividade.....	25
Figura 9 – Mapa de amplitude altimétrica.	27
Figura 10 – Mapa de curvatura.	28
Figura 11 – Mapa de radiação solar.....	29
Figura 12 – Mapa de Índice topográfico combinado.	30
Figura 13 – Mapa de unidades litológicas da área em estudo, elaborado por COMIG (1997) e CPRM (2004).	31
Figura 14 – Mapas de índices derivados das imagens ETM+: A) <i>Clay minerals</i> , B) <i>Ferrous minerals</i> , C) <i>Iron Oxide</i> e D) NDVI.	34
Figura 15 – Mapa de solos da área em estudo elaborado pelo Projeto RADAMBRASIL.....	36
Figura 16 – Localização dos perfis de solos e pontos de referência utilizados na classificação e validação do mapa de solo classificação.	40
Figura 17 – Mapa de solos obtido pelo Método convencional (Schaefer et al., 2008)..	43
Figura 19 – Valor máximo dos índices derivados da imagem para as unidades de solo: 1))PVAd, 2)PVd + PVAd, 3)CXbd1, 4)CXbd, 5)LVd + PVd, 6)LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.	50
Figura 18 – Valores máximo das variáveis altitude e amplitude altimétricas nas amostras de validação das unidades de solo: 1))PVAd, 2)PVd + PVAd, 3)CXbd1, 4)CXbd, 5)LVd + PVd, 6)LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.....	51
Figura 20 – Frequência da variável aspecto nas amostras de validação para as unidades de solo: 1)) PVAd, 2)PVd + PVAd, 3) CXbd1, 4) CXbd, 5) LVd + PVd, 6) LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.	59
Figura 21 – Mapa de solos obtido por classificação pelas Redes Neurais Artificiais...	69
Figura 22 – Mapa de solos obtido por classificação pelo MaxVer.....	70

LISTA DE QUADROS

Quadro 1 – Classes de solo identificadas no levantamento e número de perfis amostrados	20
Quadro 2 – Área das classes de aspecto	23
Quadro 3 – Área das classes de declividade	26
Quadro 4 – Unidades litológica da área em estudo	32
Quadro 5 – Características das bandas espectrais da imagem do satélite Landsat, sensor ETM + utilizadas na classificação dos solos.....	32
Quadro 6 – Unidades de solos mapeadas na área em estudo pelo Projeto RADAM BRASIL na escala de 1:1.000.000	37
Quadro 7 – Unidades de solo mapeadas pelo método convencional.....	44
Quadro 8 – Descrição das unidades de solo	47
Quadro 9 – Valores dos atributos do terreno e das amostras de validação para as unidades de solo.....	49
Quadro 10 – Caracterização estatística dos índices derivados da imagem para as unidades de solo nas amostras de validação da classificação	51
Quadro 11 – Correlação entre variáveis discriminantes para as amostras de treinamento das unidades de solos.....	52
Quadro 12 – Matriz de significância para número de neurônios na camada interna da rede	54
Quadro 13 – Matriz de significância para o número de ciclos de treinamento.....	55
Quadro 14 – Matriz de significância para as classificações feitas com 15 diferentes conjuntos de variáveis pelas redes neurais	58
Quadro 15 – Matriz de confusão para a classificação feita pelas redes neurais com todas as 19 variáveis.....	61
Quadro 16 – Matriz de confusão para a classificação feita pelas redes neurais sem a variável <i>Clay mineral</i>	62
Quadro 17 – Matriz de significância entre as classificações feitas com 15 conjuntos de variáveis discriminantes pelo MaxVer	65
Quadro 18 – Matriz de confusão gerada na classificação pelo MaxVer com 11 variáveis.....	66
Quadro 19 – Matriz de significância para as melhores classificações obtidas pelas redes neurais e pelo MaxVer no treinamento do classificador	67
Quadro 20 – Matriz de significância para a validação com pontos de referência terrestre para as classificações obtidas pelas Redes Neurais e pelo MaxVer.....	71

Quadro 21 – Matriz de significância estatística para as unidades de solo classificadas pelas Redes Neurais e pelo MaxVer	74
Quadro 22 – Valores obtidos na validação com pontos de referência terrestre para as classificações feitas pelo MaxVer e Redes Neurais	75
Quadro 23 – Concordância entre os pontos de referência de solos e os componentes de associação das unidades de mapeamento do mapa convencional.....	77
Quadro 24 – Valores de exatidão para os componentes de associação e unidades de mapeamento do mapa gerado pelo método convencional	79

LISTA DE ANEXOS

Quadro 1 – Matriz de confusão para a classificação feita pelas redes neurais sem a variável NDVI.....	87
Quadro 2 – Matriz de confusão para a classificação feita pelas redes neurais para o conjunto sem a variável aspecto.....	88
Quadro 3 – Matriz de confusão do treinamento da classificação feita pelo algoritmo MaxVer com 19 variáveis discriminantes	89
Quadro 4 – Matriz de confusão obtida no treinamento da classificação pelo MaxVer sem a variável geologia	90
Quadro 5 – Matriz de confusão da classificação pelo MaxVer sem as variáveis discriminantes geologia e curvatura	91
Quadro 6 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelas redes neurais com todas as 19 variáveis.....	92
Quadro 7 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelas redes neurais para o conjunto com 18 variáveis (sem a variável <i>Clay mineral</i>).....	93
Quadro 8 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelo MaxVer para o conjunto com 11 variáveis (sem as variáveis geologia, curvatura e bandas espectrais da imagem de satélite)	94
Quadro 9 – Matriz de confusão para as classes de solo do Primeiro Componente de associação das unidades do mapa gerado pelo método convencional	95
Quadro 10 – Matriz de confusão para as classes de solo do Segundo Componente de associação das unidades do mapa gerado pelo método convencional ...	95
Quadro 11 – Matriz de confusão para as classes de solo do Terceiro Componente de associação das unidades do mapa gerado pelo método convencional ...	95
Quadro 12 – Matriz de confusão ponderada para as classes de solo do Primeiro Componente de associação das unidades do mapa gerado pelo método convencional.....	96
Quadro 13 – Matriz de confusão ponderada para as classes de solo do Segundo Componente de associação das unidades do mapa gerado pelo método convencional.....	96
Quadro 14 – Matriz de confusão ponderada para as classes de solo do Terceiro Componente de associação das unidades do mapa gerado pelo método convencional.....	96

RESUMO

SOUZA, Eliana de, M. Sc., Universidade Federal de Viçosa, fevereiro de 2009.
Classificação supervisionada de solos por redes neurais artificiais na Serra do Cipó - MG. Orientador: Elpídio Inácio Fernandes Filho. Co-orientadores: Carlos Ernesto G. R. Schaefer e João Carlos Ker.

A classificação supervisionada de solos, especialmente nas últimas décadas, vem sendo realizada com o auxílio de modelos matemáticos e estatísticos, dentre os quais destaca-se o modelo de redes neurais, o qual tem apresentado exatidão superior quando comparado com métodos clássicos, como o de Máxima Verossimilhança (MaxVer), auxiliando no método convencional de mapeamento. No entanto, na maioria dos trabalhos foram avaliadas as propriedades dos solos, sendo o estudo das classes de solos ainda incipiente. Assim, este trabalho teve como objetivo realizar a classificação de solos por redes neurais e pelo MaxVer para uma área situada na Serra do Cipó, no estado de Minas Gerais. Para tanto, utilizaram-se informações analíticas de 55 perfis de solos, classificados até o quarto nível categórico do Sistema Brasileiro de Classificação de Solos. As unidades do mapa de solos foram compostas por semelhanças entre as propriedades físicas do solo e as características do ambiente. As variáveis discriminantes avaliadas na classificação 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: altitude, declividade, índice topográfico combinado, face de exposição, radiação solar, curvatura e amplitude altimétrica, além dos mapas geológico e pedológico. A partir desse conjunto de variáveis, identificaram-se aquelas que melhor contribuíram na discriminação dos solos, em cada uma das duas abordagens empregadas. Na classificação pelas redes neurais foram empregados o simulador *Stuttgart Neural Network Simulator* e o algoritmo *backpropagation*, sendo a arquitetura e os parâmetros selecionados por meio de tentativas e testes de significância estatística. Os resultados obtidos por ambos os classificadores, redes neurais e MaxVer, foram comparados entre si, utilizando-se a validação dos mapas com pontos de referência terrestre. Os mesmos pontos de referência foram utilizados para validar o mapa de solos obtido pelo método convencional de mapeamento. Os mapas obtidos pelos dois classificadores, utilizando o conjunto de variáveis que proporcionou melhor desempenho do classificador, apresentaram índice de exatidão considerado bom, sem diferença estatística na exatidão global dos mapas. O mapa melhor classificado pelo MaxVer apresentou índice kappa de

0,58, enquanto que, pelas redes neurais, o maior índice foi de 0,60. Esses valores não diferiram estatisticamente, entretanto, os classificadores diferiram na discriminação das unidades de solo, 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 obtido pelo método convencional de mapeamento foi de 82%, sendo esse índice calculado pelo somatório dos solos de referência concordantes com qualquer componente da unidade. Os solos no primeiro componente das unidades de mapeamento apresentaram 48% de concordância com solos de referência.

ABSTRACT

SOUZA, Eliana de, M.Sc., Universidade Federal de Viçosa, February, 2009.
Supervised classification of soils using artificial neural networks in the Serra do Cipó - MG. Adviser: Elpídio Inácio Fernandes Filho. Co-advisers: Carlos Ernesto G. R. Schaefer and João Carlos Ker.

The supervised classification of soils, especially in recent decades, is being carried out using mathematical and statistical models, amongst which the model of neural networks, stand out by greater accuracy of mapping comparing to classical models, such as Maximum Likelihood (MaxVer), helping the conventional method of mapping. Neural Networks model has been performed mostly for soil properties, with little application for soil classes. This work aimed to undertake the classification of soil using neural networks and MaxVer for an area located in Serra do Cipó, in the State of Minas Gerais. The map units was defined based on information from 55 soil profiles classified accord to the Brazilian System of Soils Classification to the fourth categorical level, whereas compound units were done with basis on the similarity of soil properties and the characteristics of physics environment. The discriminates variables used included: six scenes of Landsat satellite image sensor ETM+, four indexes derived from this image (Clay minerals, Ferrous minerals, Iron oxide and NDVI), Digital Elevation Model and derived attributes: altitude, slope, compound topographic index, aspect, solar radiation, curvature and elevation amplitude, in addition to geological and soil maps. Several variables combinations were tested in both classifiers, selecting those that best contribute to classify the soil with high accuracy on the two supervised mapping approaches. The classification by neural networks was performed using the Stuttgart Neural Network Simulator and the backpropagation algorithm, the framework and classification parameters were selected by training and statistical tests. The results obtained with both classifiers, neural networks and MaxVer were compared using ground data as reference. The same set of reference points was used to validate the soil map obtained by the conventional method of mapping. Maps obtained by the two classifiers using the group of variables that provided the best performance to the classification showed a good accuracy index, with no statistical difference in overall accuracy of the maps. The map generated by MaxVer showed a kappa index of 0.58, while the map from neural network showed an index of 0.60. Although the accuracy of the two maps was statistically similar, the classifiers efficiency in individual

discrimination of soil units differed significantly, with two units being best classified by MaxVer, three units by neural networks and four units with similar accuracies in both approach classification. The overall accuracy of soil maps made by the conventional method was 82%. The soils of the first component in mapping units agreed in 48% with reference soils.

1. INTRODUÇÃO

Os mapas de solos produzidos por meio do levantamento convencional são as principais fontes de informação pedológica para modelagem e atividades de manejo. À medida que vem aumentando a conscientização em torno dos problemas advindos da poluição ambiental e da contaminação dos solos, torna-se crescente a busca por mapas com informações quantitativas cada vez mais acuradas que permitam tomar decisões quanto ao manejo e uso dos solos.

O desenvolvimento de modelos numéricos de classificação de solos visa superar o alto custo e a morosidade da execução dos levantamentos pelo método convencional, somado à carência desses mapas em médias escalas. Além disso, os mapas produzidos pelo método convencional são estritamente qualitativos e raramente apresentam alguma estimativa da confiabilidade da informação que permita tomar decisões mais precisas acerca do uso da informação apresentada.

A demanda por informações pedológicas não oferecidas pelo modelo tradicional de levantamento e os avanços nos sistemas computacionais de informações geográficas têm levado ao desenvolvimento de modelos de mapeamento digital de solos, a exemplos dos modelos de Redes Neurais Artificiais (RNAs). No entanto, os trabalhos de classificação de classes de solos empregando RNA ainda são escassos e foram realizados para pequenas áreas, não abrangendo unidades fisiográficas representativas de muitos solos.

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. 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) para uma área na Serra do Cipó, no estado de Minas Gerais;

- 2) determinar as variáveis preditivas que melhor discriminam os solos nos modelos de classificação por RNA e MaxVer;

- 3) avaliar a eficiência dos classificadores de RNAs e do MaxVer na predição dos solos, por meio de índices de exatidão calculados com pontos de referência terrestre, e comparar as informações pedológicas obtidas na classificação por esses classificadores, com aquelas de mapas feitos pelo método convencional de mapeamento.

2. REVISÃO DE LITERATURA

2.1. Mapeamento digital de solos

Os mapas de solos constituem uma fonte primária de informações necessárias em diversos projetos de manejo ambiental, além do tradicional uso agrônomico. Atualmente, os mapas produzidos por meio do levantamento convencional são as principais fontes de informação sobre a distribuição espacial dos tipos de solos para técnicas de modelagem e atividades de manejo (ZHU, 2000).

A demanda por informações pedológicas vem, hoje, de vários segmentos da sociedade, como instituições não governamentais de planejamento do uso da terra, agricultores que empregam as informações em atividades, como produção agrícola, previsão de safras, educação ambiental, conservação e em trabalhos de engenharia (KROL, 2008). Para esses usuários, na maioria das vezes, as informações necessárias referem-se mais aos aspectos interpretativos para uso dos solos, como a aptidão das terras e as características físicas limitantes para determinados projetos de engenharia, do que ao tipo de classe de solo. Entretanto, usos mais modernos da informação pedológica demandam mapas cada vez mais detalhados e mais precisos, que apresentem dados quantitativos e uma estimativa sobre a confiabilidade das informações mapeadas (HENGL, 2003; CARRÉ & BOETTINGER, 2008; KROL, 2008).

O modelo tradicional de mapeamento de solos é feito a partir da imagem mental do pedólogo, embasada em observações da paisagem *in loco*, amostragem e análise de solos, interpretações de fotografias aéreas, de imagens de satélites e de radar, e de mapas do meio físico, como o geológico e o de relevo (ROSSITER, 2005). Nessa abordagem, é necessária uma grande habilidade do pedólogo para o delineamento das unidades de mapeamento. Além disso, o modelo de distribuição dos solos na paisagem fica restrito ao pedólogo e não pode ser reproduzido.

Segundo Hudson (1992), o modelo convencional de mapeamento é criticado por apresentar-se muito mais como arte do que como ciência. Críticas a esse modelo também recaem no alto custo com trabalho de campo, análises laboratoriais e também na morosidade do processo, principalmente, quando realizado em grande escala e em extensas áreas (McBRATNEY et al., 2003; CHAGAS, 2006). Outra crítica refere-se ao fato de o modelo apresentar informações estritamente qualitativas e, na maioria das vezes, sem uma estimativa da confiabilidade das informações apresentadas nos mapas, o

que contribui para demonstrar as falhas e limitações desse modelo (McKENZIE & RYAN, 2000; ROSSITER, 2005).

Os mapas de solos obtidos por método convencional têm sido apontados como incapazes de oferecer informações adequadas para modelagem em média escala (ZHU, 2000). Nesse tipo de modelagem, as informações requeridas devem apresentar um nível de detalhe compatível com aquela dos demais dados empregados, os quais derivam de produtos de maiores escalas e resolução, a exemplo dos modelos digitais de elevação e imagens de satélites obtidas em alta resolução (LAGACHERIE, 2008).

A disponibilidade de informação pedológica desatualizada e com exatidão desconhecida faz com que muitos estudos derivem informações a partir de levantamentos de solos, muitas vezes, defasados ou com escalas incompatíveis com os níveis de detalhes exigidos. Cumpre enfatizar também que a integração de mapas de solos pouco detalhados e de exatidão desconhecida a outras bases de dados produz resultados inconsistentes e de baixa confiabilidade (BISHOP et al., 2006).

Os mapas de solos existentes no Brasil foram, em sua maioria, obtidos pelo método convencional de levantamentos e, grande parte deles, encontra-se ainda em formato analógico, o que leva a prever que, em função do custo e do tempo necessários para a conversão de tais informações para o formato digital, talvez esse procedimento nunca seja feito por completo (McBRATNEY et al., 2003). Aliado a isso, parte desses mapas apresentam escala pequena, tornando incompatível seu uso em modelos implementados com dados de maior escala e resolução (ZHU et al., 1997).

As limitações apresentadas pelo modelo convencional de mapeamento de solos, tais como o modelo mental, baseado em polígonos com limites definidos de forma subjetiva, bem como a baixa confiabilidade das informações, o elevado custo, e o tempo gasto na execução levaram ao desenvolvendo de abordagens baseadas em técnicas digitais para estudos quantitativos dos solos, desde o início da década de 90 (HENGL, 2003; BURROUGH, et al., 1994; COOK et al., 2008).

As abordagens de mapeamento baseadas em modelos quantitativos são conhecidas na ciência do solo como pedometria (WEBSTER, 1994). Os estudos pedométricos baseiam-se em modelos quantitativos 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; HEMPEL et al., 2008; HOWELL et al., 2008).

Tais estudos estão apoiados nos avanços dos sistemas computacionais e de informações geográficas, e na disponibilidade de informações do meio físico adquiridas por sensores remotos que estão em constante desenvolvimento.

As informações quantitativas de propriedades de classes de solos vêm sendo requisitadas, sobretudo mais recentemente, devido ao aumento da conscientização em torno dos problemas advindos da poluição ambiental e da contaminação dos solos, associado ao manejo desse recurso (McBRATNEY et al., 2000). Nesse contexto, os estudos pedométricos têm sido desenvolvidos empregando-se técnicas para o mapeamento com vários níveis de detalhe e escalas (HOWELL et al., 2008).

Os trabalhos de levantamento de solos têm, no processo de coleta de amostras em campo, o maior dispêndio de recursos (McBRATNEY et al., 2003; ZHU, 2008). Os modelos de mapeamento digital de solos, realizados por meio de correlação ambiental, trazem como vantagem o poder de prever informações sobre áreas não visitadas, a partir de amostras de treinamento, com elevado grau de exatidão.

Segundo Rossister (2000), o uso de correlação ambiental necessita de menor número de amostras de solo e com isso reduz o custo final do mapeamento. Contribuem para isso o uso de ferramentas de geoprocessamento empregadas na coleta, tratamento e análise de dados, como também bancos de dados formados por imagens de satélites e dados auxiliares, os quais vêm sendo disponibilizados gratuitamente na internet ou vendidos a custos acessíveis à maioria dos usuários (HOWELL, 2008).

Os mapeamentos digitais de solos têm grande potencial de aplicação, sobretudo, no levantamento dos solos de áreas ainda não mapeadas ou com base de dados escassa e pouco detalhada (HEMPEL et al., 2008). Em áreas já mapeadas, surge a necessidade de atualização do mapeamento, em razão das mudanças no sistema de classificação dos solos e mesmo nas propriedades destes (McBRATNEY et al., 2003).

Os trabalhos prioritários no mapeamento digital de solos visam ao desenvolvimento de métodos e técnicas que contribuam para a caracterização da confiabilidade das informações mapeadas (COOK et al., 2008), pois nos processos de tomada de decisão sobre o uso de mapas de solos são necessárias medidas estatísticas sobre a qualidade das informações.

A estimativa das incertezas dos mapas visa eliminar a propagação do erro no processo de modelagem e tem sido requisitada pelos usuários que precisam decidir sobre o uso de tais mapas em seus modelos. De acordo com Burrough e McDonnell (2004), em razão do uso de ferramentas automatizadas empregadas no mapeamento é

crescente o número de mapas de solos que apresentam alguma estimativa da confiabilidade da informação.

Vários são os modelos utilizados na predição de informações pedológicas, a exemplo dos estatísticos, geoestatísticos, árvore de decisão, regressão linear e sistemas inteligentes. Um modelo que tem sido utilizado, ainda de forma incipiente, é o das Redes Neurais Artificiais (McBRATNEY et al., 2002). Esses modelos têm sido empregados principalmente para mapear propriedades dos solos, tais como profundidade, pH, CTC, densidade, textura e conteúdo de carbono orgânico.

Mcbratney et al. (2003) realizaram um levantamento dos trabalhos de predição de propriedades e/ou classes de solos, no qual se pode observar que o mapeamento quantitativo de informações pedológicas vem sendo realizado desde 1964, ocorrendo intensificação do número de trabalhos de modelagens nas duas últimas décadas. As informações obtidas em tais trabalhos têm sido apresentadas e aceitas pela comunidade científica. Os estudos avaliam a adoção de modelos que possam ser usados como padrão para mapeamento digital de solos (HOWELL et al., 2008).

As pesquisas prioritárias em mapeamento de solos estão focadas em estudos dos dados utilizados, em sistemas de predição dos solos e em métodos para mensurar a qualidade dos dados e informações geradas (CARRÉ & BOETTINGER, 2008). Esses estudos visam definir protocolos e garantir padrões de qualidade das informações geradas, de modo que os usuários possam decidir sobre o uso das mesmas.

Howell et al. (2008) acreditam que um desafio estrutural ao emprego do mapeamento digital de solos, tanto de propriedades quanto de classes, passa pela adoção e comunicação entre instituições reguladoras dos levantamentos de solos, de um modelo padrão que permita o seu desenvolvimento e aplicação. Outro desafio refere-se ao desenvolvimento de métodos para interpretar as informações obtidas.

O desenvolvimento dos modelos quantitativos de mapeamento de solos é grandemente dependente da incorporação do conhecimento e da experiência dos pedólogos em parâmetros de funcionamento dos modelos e da interpretação dos resultados obtidos (BUI, 2004; HENGL, 2003).

O desenvolvimento dos modelos de classificação tem como desafio a dificuldade no mapeamento de solos que apresentam horizontes com transição abrupta e solos de rochas calcárias, os quais são ainda melhor separados utilizando-se o método de mapeamento convencional (HENGL, 2003). Outro aspecto relaciona-se com o estado disperso em que se encontra a literatura dos estudos pedométricos. Neste sentido existe

a necessidade de se organizar e sistematizar o conhecimento relacionado aos estudos pedométricos para que os usuários possam fazer escolhas sobre as incertezas dos produtos empregados em modelagens (BURROUGH et al., 1994).

Assim, a integração do conhecimento mental dos solos, adquirido pelos pedólogos, às técnicas utilizadas nos modelos pedométricos visa tornar o mapeamento mais científico e quantitativo, conciliando as técnicas do mapeamento convencional às da pedometria.

2.2. Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são modelos de classificação que utilizam técnicas matemáticas inspiradas no funcionamento da aprendizagem do cérebro humano para aprender e prever padrões (HAYKIN, 2001). Uma RNA é constituída por uma coleção de unidades computacionais, denominadas neurônios, conectadas de acordo com uma arquitetura específica, como apresentada na figura 1.

Na arquitetura das redes neurais, a camada de entrada representa as variáveis discriminantes dos solos e se conecta à camada de saída por uma ou mais camadas internas; a camada de saída é representada pelas classes de solo, enquanto que as camadas internas correspondem às unidades de processamento do classificador.

O número de neurônios das camadas de entrada e saída são fixos e determinado de acordo com o número de variáveis discriminantes e o total de classes do mapeamento. O número de neurônios das camadas internas, bem como o número de camadas são definidos durante o treinamento do classificador.

Uma importante característica do classificador de redes neurais é a flexibilidade estatística (MATHER, 2004). O método é não-paramétrico, o que significa que o reconhecimento dos padrões é feito sem considerar, *a priori*, a distribuição estatística dos dados. Assim, não ocorre a pressuposição de distribuição normal da frequência dos dados, como é feita pelos classificadores estatísticos clássicos como o MaxVer.

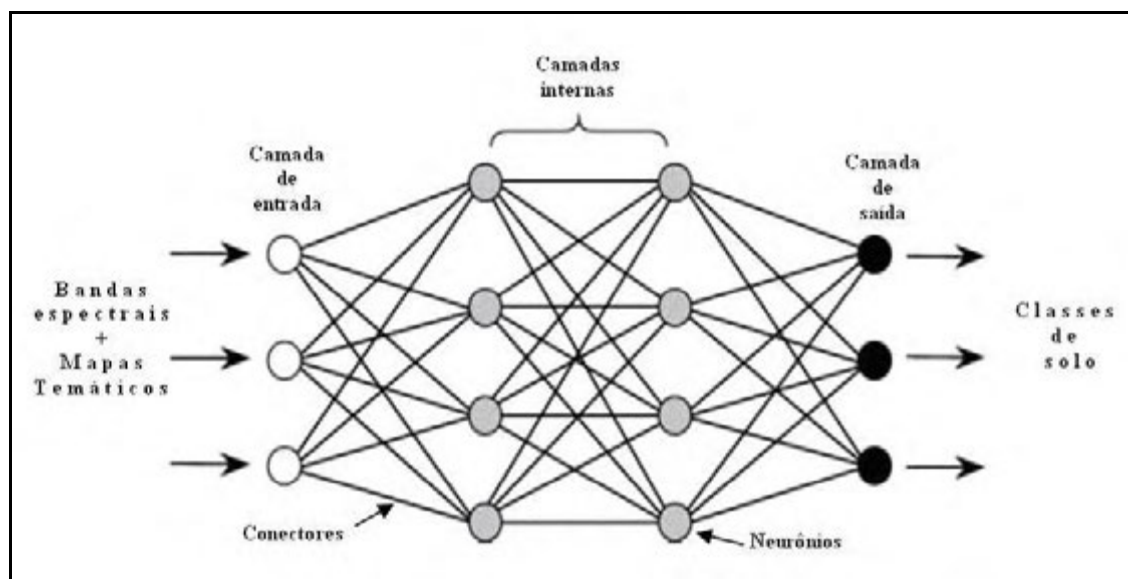


Figura 1 – Arquitetura de rede neural. Modificada de Kavzoglu e Mather (2003).

A característica não-paramétrica permite o uso de dados de múltiplas fontes e vários formatos, podendo ser tanto contínuos, como os dados de imagens de satélites, quanto categóricos, a exemplo dos dados geológicos e de vegetação. A integração de informações permite mapear propriedades graduais ou descontínuas dos solos, além de aumentar a exatidão da classificação (MATHER, 2004).

O classificador de redes neurais tem a habilidade para classificar de forma mais exata do que classificadores estatísticos, principalmente quando as classes apresentam uma distribuição complexa no espaço, e os dados discriminantes possuem diferente distribuição estatística (ATKINSON & TATNALL, 1997).

Outras vantagens de se utilizarem as redes neurais referem-se à estrutura de processamento e à tolerância a dados falhos (VIEIRA, 2000). A estrutura paralela de processamento faz com que a classificação ocorra de forma rápida, uma vez definida a arquitetura interna das redes. A tolerância a dados falhos está relacionada com o caráter robusto da RNA, o qual permite agregar dados com diferentes níveis de exatidão.

Os parâmetros básicos de funcionamento das redes neurais, como o formato de entrada dos dados, o algoritmo de aprendizagem, as funções de ativação e a arquitetura das redes foram objetos de várias pesquisas que procuraram estabelecer a forma de se obterem os melhores resultados numa classificação, de modo que o uso do classificador de redes neurais está bem estabelecido (KANNELLOPOULOS & WILKINSON, 1997).

O algoritmo *backpropagation* é o mais empregado para o treinamento do classificador na aprendizagem do reconhecimento de padrões (VIEIRA, 2000). Em

relação à definição da estrutura das redes neurais, existem algumas propostas de forma experimental para se obter um resultado razoável numa classificação. Contudo, o melhor resultado para cada caso é definido por tentativas e erros, a partir do treinamento do classificador, uma vez que cada área apresenta certa variação na complexidade de distribuição das classes mapeadas (KAVZOGLU & MATHER, 2003).

As redes neurais vêm sendo utilizadas para a classificação de imagens de sensoriamento remoto desde o final da década de 80, sendo os primeiros trabalhos realizados para a classificação de padrões de vegetação (KANNELLOPOULOS & WILKINSON, 1997). Os trabalhos de mapeamento de solos são mais recentes e foram realizados principalmente para propriedades de solos, sendo escassos aqueles para discriminar classes de solos.

Alguns estudos comparativos mostram que os classificadores de RNA podem classificar dados de forma mais exata do que outros classificadores (FOODY & ARORA, 1997). Resultados obtidos por Zhu (2000) mostram uma grande eficiência do classificador, quando comparados com o mapeamento convencional.

Carvalho-Júnior (2005) utilizou redes neurais no mapeamento de unidades de solo-paisagem, obtendo exatidão superior a 66% em todas as predições realizadas para solos de uma área de 47.292 ha, no estado do Rio de Janeiro. Todas as classificações realizadas foram consideradas confiáveis, tendo como referência validação feita com perfis de solos levantados na área pela Embrapa solos.

Zhu (2000), utilizando uma abordagem por redes neurais, mapeou solos classificados até o nível de série em uma área de 4 mil ha, para a qual a validação do mapeamento com pontos de referência coletados na área mostrou uma inferência correta de 77% para o mapa obtido pela classificação automática, ao passo que com o método convencional obteve-se 60,9% de exatidão. Nesse trabalho foi também observado que além de mais exato, o mapa classificado pelas redes neurais apresentou maior detalhe espacial do que o obtido pelo método convencional

Fidêncio et al. (2001) compararam dois modelos de RNA para classificar Latossolos e Argissolos de diferentes regiões do estado de São Paulo. Os resultados alcançados foram melhores quando se utilizou o modelo de rede *counterpropagation* neural network, tendo 91,4% de exatidão na classificação.

Chagas (2006), em estudo feito para uma bacia hidrográfica localizada no estado do Rio de Janeiro, comparou os mapas obtidos por neurais e por método convencional, através de análise de concordância entre os mapas e pontos referência. Os resultados

mostraram que a exatidão da classificação pelas redes foi de 70,8%, e de 52,8% para o primeiro nível categórico das classes de solo do mapa convencional.

Sirtoli (2008) empregou o classificador de redes neurais para a classificação de solo numa área de 303,4 ha no estado do Paraná. O mapeamento utilizou variáveis discriminantes derivadas de base de dados composta por imagem de satélite de alta resolução, atributos do terreno derivados de modelo digital de elevação com 5 m de resolução e mapa geológico. Os resultados da comparação entre o mapa obtido na classificação e um mapa feito por método convencional mostraram que as redes neurais foram eficientes para discriminar os solos mostrando mais detalhe espacial do que o apresentado no mapa gerado por método convencional de mapeamento.

2.3. Variáveis discriminantes dos solos

A seleção das variáveis para o mapeamento de solos é realizada tendo como base os fatores de formação dos solos definidos na função clássica, proposta por Jenny (1941) e modificada por McBratney et al. (2003), como o modelo *Scorpan*, apresentado a seguir:

$$Sc: f(s, c, o, r, p, a, n)$$

Em que: (Sc) corresponde à informação pedológica mapeada em função de: (s) = propriedade do solo, (c) = clima, (o) = organismo, (r) = relevo, (p) = material de origem, (a) = tempo e (n) = posição espacial.

As características fisiográficas de cada ambiente condicionam a pedogênese diferenciada, de modo que as variáveis que melhor discriminam os solos de uma área devem ser determinadas, *a priori*, conforme o modelo de mapeamento a ser empregado.

As variáveis discriminantes utilizadas nos modelos preditivos de solos compõem um conjunto de dados que pode agregar informações de solos já existentes e/ou mapeados, dados obtidos por satélites ou radar e mapas de temas como geologia, vegetação e geomorfologia. A disponibilidade de tais dados e as características diferenciadas de um local para outro levam a uma combinação variada de informações utilizadas nos diversos modelos de classificação supervisionada de solos.

Os dados de solo já levantado, de perfis ou mapas de solo têm apresentado contribuições significativas no mapeamento de solos quando utilizados em associação

com dados de imagem de satélites e atributos do terreno (VOLTZ & WEBSTER, 1990; ZHU et al, 2000, SCULL et al 2005; KROL, 2006; CARRÉ & BOETTINGER, 2008).

Os dados espectrais de imagens de satélites, muito utilizados no mapeamento geológico através dos níveis de cinza ou por meio de índices derivados (LILLESAND & KIEFER, 1994; VINCENT, 1997, RAJESH, 2004; MOREIRA, 2005; CHELLAIAH, 2003), vêm sendo empregados nos estudos pedológico para mapear propriedades e classes de solo. Os níveis de cinza da imagem podem ser entendidos como a resposta espectral do solo que varia de acordo com as características de seus constituintes, permitindo diferenciá-los por meio de propriedades como conteúdo de água, teor de matéria orgânica e tipo de minerais. A faixa do espectro do visível, do infravermelho médio e próximo são as mais representativas para a coleta de informações para o mapeamento dos solos, por meio da identificação e mapeamento dessas características.

Alguns índices derivados das imagens de satélite que se destacam entre aqueles empregados na classificação de solos refere-se ao índice de vegetação por diferença normalizada (NDVI), comumente utilizado para ressaltar o comportamento dos solos em relação à vegetação, e os índices *Clay minerals*, *Ferrous minerals* e *iron oxide*, que permitem diferenciar os solos a partir de suas características físicas e minerais (MATHER, 2004; RAJESH, 2004; DEMATTÊ et al, 2007). O NDVI apresentou contribuições significativas quando empregado por Moran e Bui (2002) e Dobos et al. (2000) no mapeamento de solo em escala regional.

Conforme Rajesh (2004), a vegetação mascara a resposta espectral do solo, de modo que, em áreas com cobertura vegetal, devido ao comportamento do solo em relação aos óxidos de ferro e minerais de argilas, a resposta espectral da vegetação se confunde com a resposta desses minerais pela similaridade de padrões.

A separação dos solos em relação ao material de origem, a partir do conteúdo dos minerais, pode ser feita utilizando-se índices espectrais derivados de imagens de satélites (*Clay minerals*, *iron oxide* e *Ferrous minerals*). O índice *Clay minerals* é utilizado para diferenciar os solos quanto ao conteúdo de minerais de argila dos minerais silicatados. O índice *Iron oxide*, para evidenciar a concentração do conteúdo de óxidos de ferro, principalmente hematita e goethita. Já o índice *Ferrous minerals* resalta o comportamento dos solos pela diferença no conteúdo de minerais que possuem ferro na composição (VINCENT, 1997; RAJESH, 2004).

A predominância da hematita é observada na faixa espectral de 0,35 – 1,5 μm , na qual é observado que esse óxido absorve a maior parte do comprimento de onda

incidido. Já a goethita mostra predominância a partir 0,9 μm . Os minerais de argila (*clay minerals*) são melhores identificados na faixa do infravermelho médio, 1,65 a 2,2 μm , na qual ocorre a maior reflectância desses, sendo identificadas as micas, sulfatos, carbonatos entre outros.

Os atributos do terreno, derivados de modelos digitais de elevação, tais como declividade, curvatura, aspecto, radiação solar e índice topográfico combinado, além da altitude, que é a informação base do modelo digital de elevação são largamente empregados nos modelos de mapeamento de informações pedológicas (ODEH et al., 1991; MOORE et al., 1993; GESSLER et al., 1995). Uma lista de trabalhos de mapeamento pedométricos com as variáveis empregadas na predição dos solos pode ser encontrada em McBratney et al. (2003).

O Índice Topográfico Combinado (CTI - *Compound Topographic Index*) é um índice hidrológico utilizado para prever a distribuição espacial de diversas propriedades do solo relacionadas ao conteúdo de água presente no solo (MOORE et al., 1993; GESSLER et al., 1995).

O CTI foi empregado no mapeamento do conteúdo de água no solo em áreas com solos saturados por Moore et al. (1993). Os autores encontraram alta correlação entre o CTI e os atributos do terreno, concluindo que, mais da metade da variabilidade das propriedades do solo, como pH, profundidade do horizonte, conteúdo de carbono orgânico, silte e argila, foram explicadas pela associação entre a declividade e o CTI.

Gessler et al. (2000), no mapeamento da profundidade e do conteúdo de carbono do solo, encontraram correlação positiva entre o CTI e a profundidade do solo. Os autores verificaram que o índice explicou em 33% a variação da profundidade do solo. O índice foi também empregado no mapeamento de paisagens e classes de solo, sendo verificada a contribuição desse índice por Moore et al., (1993), Gessler et al. (2000) Chagas (2006) e Zhu et al. (2008)

A escolha de variáveis adequadas para um mapeamento deve ser feita de forma criteriosa quando se utiliza o modelo de redes neurais. Foody e Arora (1997) argumentam que, embora a rede neural não sofra com o aumento na complexidade representada pelo volume de dados, o uso de variáveis que não contribuem de forma significativa com o processo de discriminação das classes em estudo não implica em ganho na exatidão do classificador.

Métodos analíticos, a exemplo da análise de componentes principais (CARRÉ et al., 2008) e análise de regressão (MOORE et al., 1995) são utilizados para a escolha de

variáveis discriminantes empregadas em modelos preditivos. Carré et al. (2008), utilizando a análise de componentes principais, verificaram que a combinação das informações dos mapas de altitude e índice topográfico explicavam 80% da variabilidade das unidades de paisagens estudadas, o que possibilitou eliminar do modelo utilizado informações como declividade e curvatura.

2.4. Avaliação da exatidão do mapeamento e comparação entre mapas

A qualidade de um mapa é definida pela capacidade que ele possui de prever a informação num determinado ponto do espaço (MACMILLAN, 2008). A avaliação desta é feita através da estimativa de incerteza da informação mapeada.

A visão geral sobre erros e incertezas apresentadas pelos dados e produtos geográficos é de que a tomada de consciência da existência dos mesmos permite conhecer como eles ocorrem, como podem ser gerenciados e, possivelmente, reduzidos (BURROUGH & McDONNELL, 2004).

O uso de mapas sem a análise e correção dos erros tem levado à propagação e acúmulo dos mesmos nos produtos gerados, principalmente em modelagens com banco de dados geográficos (BISHOP et al., 2006). De acordo com Krol (2008), existe uma falsa concepção de que toda a informação que se encontra no formato digital está correta, concorrendo para a propagação de erro nos mapeamentos, uma vez que a falsa crença leva a não análise e correção dos dados antes de se iniciar os trabalhos.

Os métodos para mensurar a confiabilidade de mapas obtidos por classificação digital, gerados a partir de dados de sensores remotos e geográficos têm focado principalmente a informação final do mapeamento. A exatidão temática, que representa a concordância entre a informação do mapa e aquela definida como de referência, é obtida de forma direta. Os dados de referência são normalmente coletados no campo ou a partir de fotos aéreas (ROSSITER, 2001).

A estatística multivariada calculada por meio da análise da matriz de confusão é considerada pela comunidade científica a técnica mais importante para avaliar a exatidão dos mapeamentos digitais (VIEIRA, 2000). Essa técnica permite comparar mapas obtidos por diferentes conjuntos de dados e métodos de classificação empregados para determinada área. Na matriz de confusão aparecem o índice kappa e a exatidão global. A partir das informações dessa matriz podem também derivar os coeficientes de

exatidão do produtor e do usuário. Esses índices de exatidão são detalhadamente descritos por Congalton e Green (1999).

A exatidão de um mapa depende de um conjunto de fatores, tais como: a escala e a natureza dos dados, as técnicas e os procedimentos empregados, a densidades de observações e a complexidade das classes mapeadas. Esses fatores fazem com que um valor de exatidão possa ser considerado bom para um mapeamento, mas não para outro. Por isso, a exatidão do mapa deve ser avaliada, considerando-se não apenas as informações mapeadas, mas também a qualidade e os métodos empregados na obtenção dos dados utilizados, e o processo de produção do mapa.

As normas para a produção de mapas esclarecem que mapas de solos em média escala devem apresentar em torno de 80% de exatidão na composição e pureza das unidades (EMBRAPA, 1997). Segundo Borrough (1998), tais normas têm baixa confiabilidade, uma vez que as associações de solos implicam em variações na legenda dos mapas e mudanças na concordância entre a informação de referência e a mapeada.

A avaliação da classificação durante o treinamento das redes neurais, por meio do Teste Z, determina se dois ou mais valores independentes de kappa apresentam diferença estatística significativa a um determinado nível de confiança. Esse teste permite comparar dois ou mais classificadores, ou conjuntos de variáveis, tornando possível distinguir qual deles contribui para o melhor desempenho do classificador e, por conseguinte, produz maior exatidão (CONGALTON & GREEN, 1999).

A comparação entre mapas permite decidir qual é o mais indicado para um uso específico, a partir da exatidão temática obtida por modelos preditivos ou conjuntos de dados empregados no mapeamento (BOOTS & CSILLAG, 2006). Na tomada de decisão, para escolher os modelos e os dados, os mapas gerados para uma determinada área são avaliados por meio dos mesmos índices de exatidão (STEHMAN, 2006).

Os elementos de análises dos estudos comparativos entre mapas categóricos estão relacionados com a composição e a configuração da informação. A composição refere-se às características espaciais das categorias e a configuração diz respeito à distribuição espacial destas (BOOTS & CSILLAG, 2006).

Chagas (2005) enfatiza a negligência na comparação entre a exatidão individual das classes mapeadas por diferentes modelos de classificação. As categorias do mapa podem apresentar diferenças estatísticas significativas na exatidão, como resultado de diferenças nos métodos matemáticos empregados, que não utilizam as informações discriminantes da mesma maneira.

3. MATERIAL E MÉTODOS

As etapas para a execução desse trabalho são apresentadas de forma resumida no fluxograma da figura 2, e de forma detalhada a seguir.

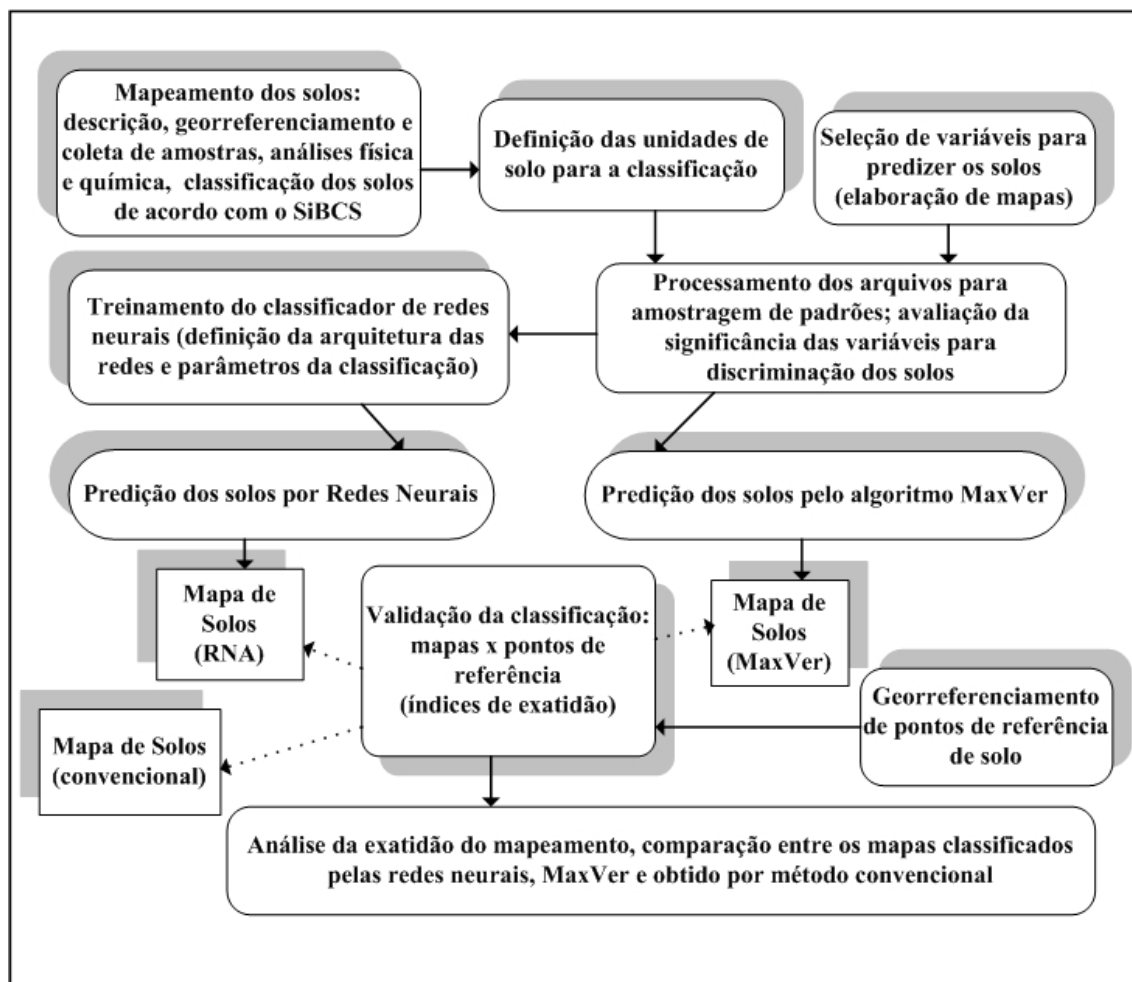


Figura 2 – Fluxograma das etapas de execução do trabalho.

3.1. Localização e caracterização da área de estudo

A área de estudo compreende a porção da Bacia Hidrográfica do Rio São Francisco inserida nas Unidades de Conservação (UC's) Parque Nacional da Serra do Cipó (PNSCi) 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. Uma área com extensão de 1 km de raio a partir do limite da APA foi incluída no estudo, obtendo uma área total de 100.716,13 ha. A área abrange parte dos municípios de Santana do Riacho, Jaboticatubas, Nova União e Taquaraçu de Minas (Figura 3).

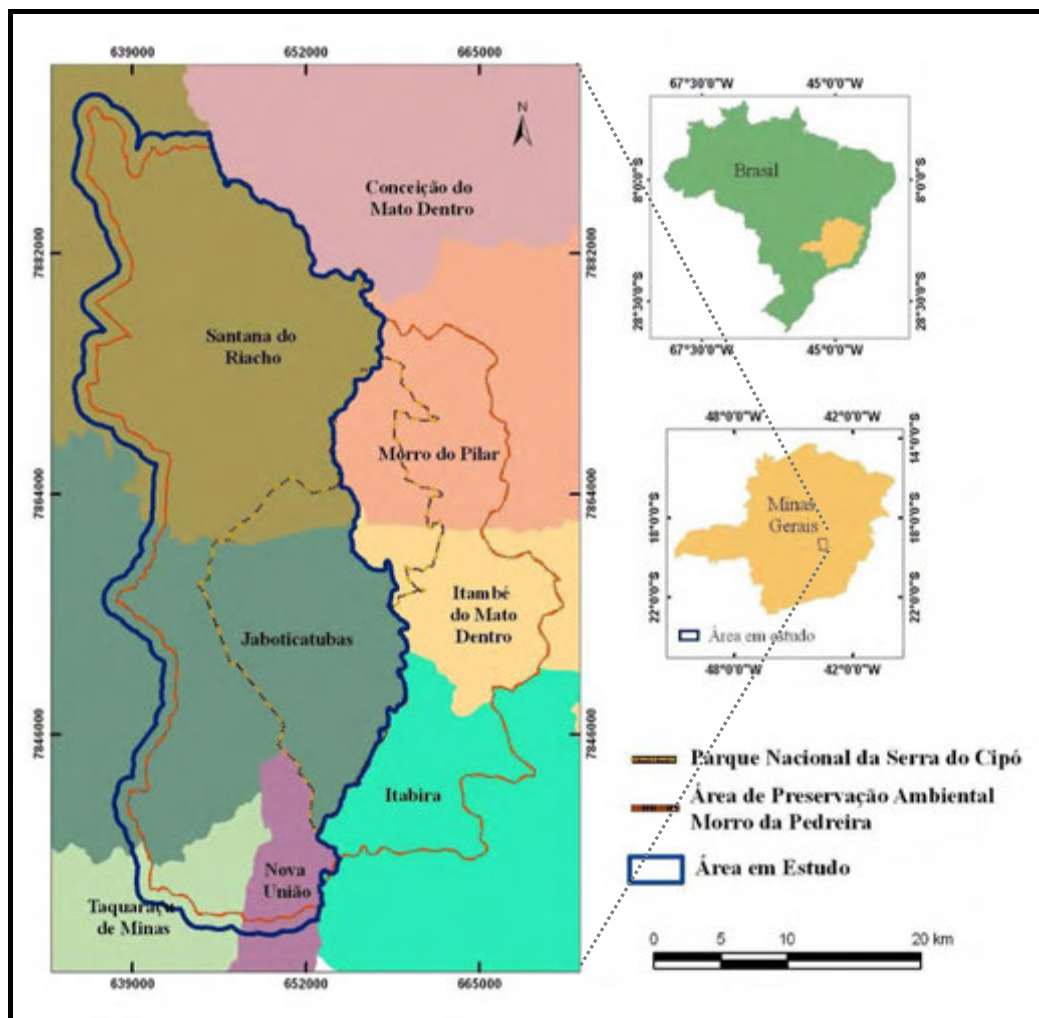


Figura 3 – Mapa de localização da área de estudo.

As UC's foram criadas para proteger o conjunto paisagístico de parte do Maciço do Espinhaço, os mananciais e o rico acervo arqueológico, com destaque para o Sítio Arqueológico de Santana do Riacho e os inúmeros registros de pinturas rupestres em abrigos e cavernas (IBAMA, 2006; IEF/MG, 2006).

Os estudos dos solos nessa área apresentam grande importância para o manejo e a conservação de importantes nascentes de afluentes do Rio São Francisco e espécies endêmicas da flora e da fauna, especialmente no domínio de Campos Rupestres.

Os eventos tectônicos que modelaram o relevo do Espinhaço produziram dobramentos com um padrão geomorfológico de expressivo contraste com as áreas adjacentes (IBAMA, 2006). Na área, o relevo pode ser dividido em dois macro-compartimentos compostos por formas, variando de plana, ondulada, montanhosa a

escarpada. Formando o alinhamento do Espinhaço, um conjunto de cristas alinhadas no sentido aproximadamente norte/sul ocorrem em altitudes que variam de 1.100 e 1.700 metros. Esse compartimento divide as duas maiores bacias hidrográficas do estado de Minas Gerais: a do Rio São Francisco e a do Rio Doce. O outro compartimento, que constitui parte da bacia Inter-planáltica do Médio Rio Cipó e está em altitudes variantes entre 700 e 900 m, forma os vales.

A geologia é composta por rochas do Super-Grupo Espinhaço, Grupos Macaúbas e Bambuí e por Depósitos Quaternários, formando um mosaico de rochas com predomínio das metamórficas com variados graus de metamorfismo. Dentre as unidades litológicas, destacam-se: rochas quartzíticas e metapelíticas, com recobrimento de mais de 50% da área. Existem também as cristalinas granítico-gnaissicas, calcárias e, em menor extensão, as metabásicas e os sedimentos (COMIG, 1997).

A vegetação é composta por três domínios fisionômicos com predominância de Campo Rupestre, na faixa central da área, com alinhamento aproximadamente norte sul.

A área está inserida numa zona ecotonal entre os Biomas Mata Atlântica e Cerrado. A divisão entre esses dois Biomas é feita por uma extensa faixa de vegetação de Campo Rupestre. A oeste predominam vegetações de Cerrado, estratificadas entre as tipologias de Cerradão, Cerrado *stricto sensu* e Campo Cerrado. A faixa formada pelo alinhamento do Espinhaço, centro/leste, é coberta por Campo Rupestre. No extremo sul da área, formando vales sob rochas granítico-gnaissicas ocorrem florestas com espécies da Mata Atlântica (IBAMA, 2006, SCHAEFER et al, 2008).

Os solos da área, mapeados na escala 1:100.000 por Schaefer et al. (2008), representam os geoambientes típicos dos três domínios da vegetação local, quais sejam: Cerrados, Campos Rupestres e Mata Atlântica.

No domínio dos Campos Rupestres, os complexos de solos formados por associações de Neossolos Litólicos, Cambissolos e afloramento de rochas com predominância dos primeiros abrangem as áreas de maiores altitudes, compostas por rochas quartzíticas. Nos sítios cujo relevo é plano, são encontrados Neossolos Quartzarênicos, Espossolos e Cambissolos húmicos.

Os Latossolos Vermelhos e Vermelho-Amarelos aparecem em áreas de rochas cristalinas e biotita-xisto, sob vegetação de Mata Atlântica. Na borda oeste da área predominam Latossolos Vermelho-Amarelos e Cambissolos, sob vegetação de Cerrados e substrato composto por rochas metapelíticas. Os Gleissolos, com menor expressão geográfica, ocorrem em áreas ao longo da rede de drenagem.

3.2. Base de Dados

A base de dados cartográficos foi implementada no software de Sistema de Informação Geográfico ArcGIS 9.2 (ESRI, 2007), a partir da compilação das seguintes informações: 1- limites territoriais das UC's (IBAMA, 2006a), 2- cartas topográficas: folha Jaboticatubas, na escala 1:50.000, e folhas Baldim, Conceição do Mato Dentro e Itabira, na escala 1:100.000 (IBGE, 1997), 3- mapas geológicos: folha Baldim e Conceição do Mato Dentro (COMIG, 1997), na escala 1:100.000; mapa geológico da CPRM (2004), na escala 1:1.000.000, 4- imagens do sensor ETM+ do satélite Landsat 7 do ano de 2006, 5- mapa de solos das UC's, elaborado pelo Plano de Manejo, escala 1:100.000 (SCHAEFER et al., 2008) e 6- mapa de solos do Projeto RADAMBRASIL Folha Belo Horizonte - SE 23, na escala de 1: 1000.000 (não publicado).

No software ArcGIS 9.2 (ESRI, 2006) foram realizados: o processamento dos dados cartográficos, a correção de falhas e redução da dimensionalidade da base de dados e a elaboração dos mapas de atributos do terreno.

O software de processamento de imagem ERDAS IMAGINE 8.5 (2001) foi empregado no processamento da imagem de satélite, na elaboração dos índices derivados desta e em etapas do processo de classificação.

A classificação foi realizada utilizando-se o simulador de redes neurais JAVA NEURAL NETWORK SIMULATOR. Os arquivos de dados utilizados na classificação foram manipulados empregando-se os aplicativos *Funcpow* e *Gerapat* (VIEIRA, 2000).

O controle de qualidade da base de dados geográficos foi feito segundo normas e critérios cartográficos definidos para as informações geográficas utilizadas em SIG (BURROUGH & McDONNELL, 2004; MATHER, 2004; KROL, 2008). Foram corrigidos e analisados parâmetros cartográficos, como precisão de posicionamento e de atributos, completude, sistemas de projeção e *Datum*, compatibilidade de escala e resolução, atualidade da informação e os algoritmos para gerar e derivar mapas. Uma vez corrigida as incertezas das informações da base de dados, a atenção foi dada à análise das variáveis discriminantes.

3.3. Dados de Solos

O mapeamento dos solos seguiu as normas estabelecidas para o levantamento pedológico na escala de 1:100.000, conforme EMBRAPA (1995), resumidamente apresentadas a seguir:

- Coletas de solos realizadas com frequência de amostragem de um perfil completo e um ponto de amostra por classe de solo em unidades simples ou componentes de associação;
- Área mínima mapeável de 10 hectares;
- Unidades de mapeamento constituídas por unidades simples ou por associações de até três componentes de subgrupos de solos;
- Imagens de satélite e cartas topográficas em escala igual ou maior que 1:100.000.

As informações pedológicas utilizadas neste estudo foram obtidas durante a elaboração do Plano de Manejo das Unidades de Conservação da área estudada para o levantamento de solos, em escala de 1:100.000 (SCHAEFER et al., 2008). Foram reconhecidos e coletados solos de 55 perfis (Quadro 1), classificados até o quarto nível categórico do Sistema Brasileiro de Classificação (EMBRAPA, 2006). A localização dos perfis estudados é mostrada na figura 4.

A área em estudo foi percorrida, buscando acessar todos os locais que apresentassem variabilidade de solos para aquisição de pontos de referência para validar o mapeamento. Para tanto, 396 pontos de referência de solos foram identificados e georreferenciados com sistema de posicionamento global (GPS).

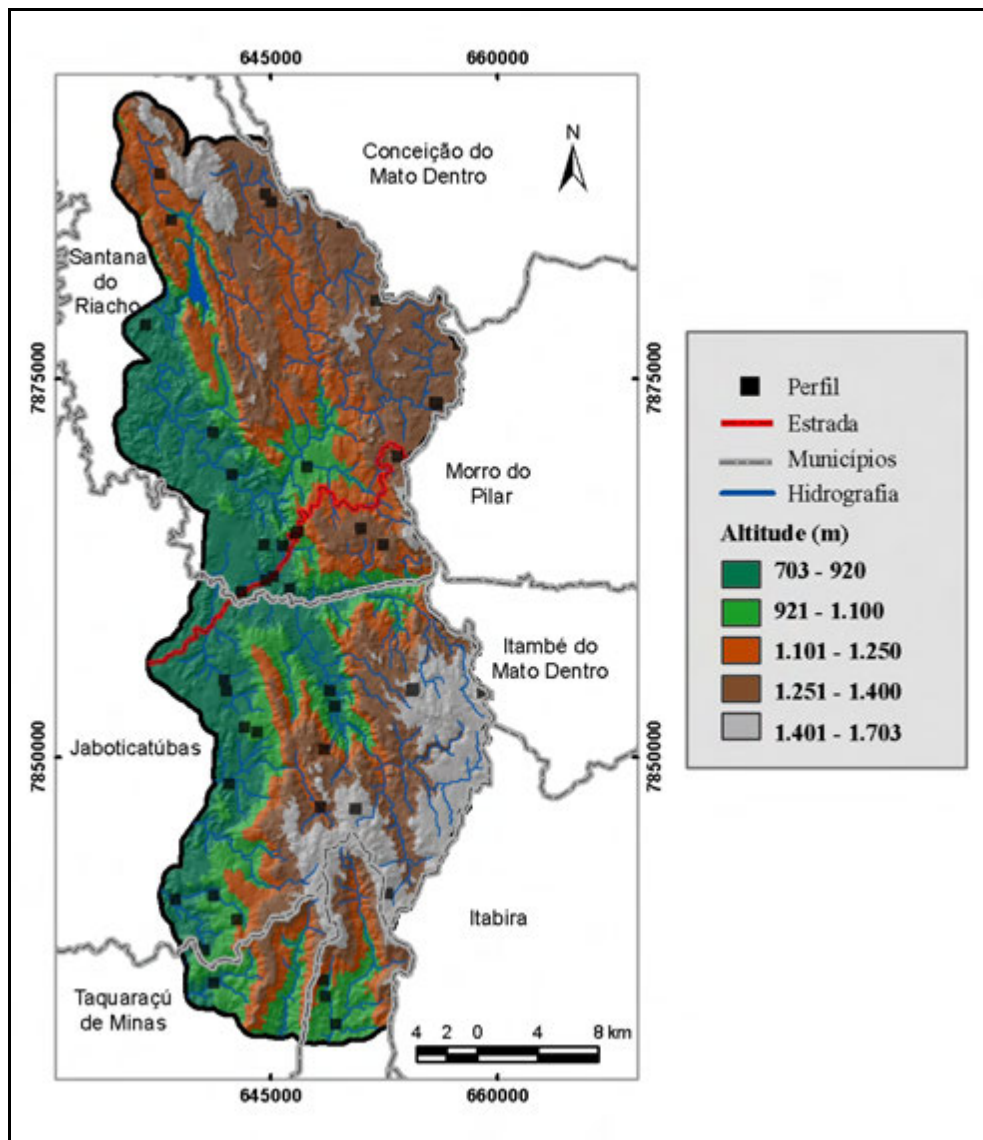


Figura 4 – Mapa de localização dos perfis de solos utilizados.

Os Cambissolos representaram a maioria dos perfis, totalizando 19 pontos amostrados; em seguida, os Latossolos e Neossolos, com 13 perfis cada classe; Argissolos, Espodossolos e Organossolo foram identificados em 3 perfis, cada classe; a classe dos Gleissolos foi identificada em apenas um dos perfis.

Quadro 1 – Classes de solo identificadas no levantamento e número de perfis amostrados

Classe de solo até o 4º nível categórico	Ordem	Nº de perfil
Argissolo Vermelho-Amarelo Distrófico típico	ARGISSOLOS	3
Argissolo vermelho Distrófico típico		
Cambissolo Háptico Tb Distrófico latossólico	CAMBISSOLOS	19
Cambissolo Háptico Tb Distrófico léptico		
Cambissolo Háptico Tb Distrófico típico		
Cambissolo Húmico Distrófico léptico		
Cambissolo Flúvico Tb Distrófico Gleissólico		
Espodossolo Ferrihumilúvico Órtico arênico	ESPODOSSOLOS	3
Espodossolo Humilúvico Órtico arênico		
Gleissolo Háptico Tb Distrófico típico	GLEISSOLO	1
Latossolo Amarelo Distrófico típico	LATOSSOLOS	13
Latossolo Amarelo Distrófico câmbico		
Latossolo Vermelho-Amarelo Distrófico típico		
Latossolo Vermelho Distrófico típico		
Neossolo Flúvico Psamítico	NEOSSOLOS	13
Neossolo Litólico Distrófico típico		
Neossolo Litólico Húmico Espódico		
Neossolo Litólico Húmico típico		
Neossolo Litólico Psamítico típico		
Neossolo Quartzarênico Órtico espódico		
Neossolo Quartzarênico Órtico típico		
Neossolo Regolítico Distrófico léptico		
Organossolo Fólico Hémico típico	ORGANOSSOLOS	3
Organossolo Háptico Sáprico típico		

3.4. Variáveis discriminantes dos solos

As variáveis testadas na discriminação dos solos foram escolhidas com base na revisão de literatura de trabalhos de predição de propriedades e/ou classes de solo, e indicação de especialista em solos com conhecimento na área.

As variáveis discriminantes avaliadas na classificação dos solos foram: geologia, altitude – representada pelo Modelo Digital de Elevação (MDE) –, atributos derivados do MDE (aspecto, curvatura, amplitude altimétrica, declividade, índice topográfico combinado e radiação solar), seis bandas espectrais do sensor ETM+, bandas do visível, infravermelho próximo e infravermelho médio, quatro índices derivados dos dados espectrais: *Clay minerals*, *Ferrous minerals*, *Iron oxide* e *NDVI*, e informação pedológica do mapa de solos (Figura 5).

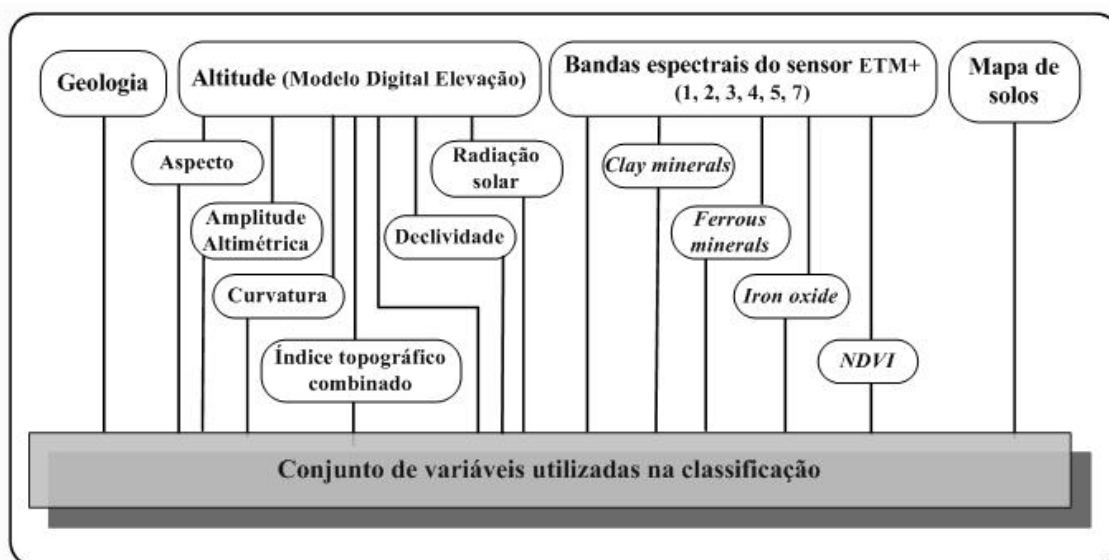


Figura 5 – Fluxograma das variáveis discriminantes utilizadas.

3.4.1. Modelo Digital de Elevação

O Modelo Digital de Elevação (MDE) foi obtido a partir da interpolação de curvas de nível extraídas das cartas topográficas do IBGE. A interpolação foi feita utilizando-se o algoritmo *Topo to Raster*, disponível no ArcGis 9.2. O modelo foi gerado com células de 30 m. Posteriormente foi realizada eliminação das depressões espúrias (células com valor inconsistentes), introduzidas no modelo durante o processo de interpolação, gerando modelo hidrológicamente consistente.

O MDE foi utilizado para derivar variáveis secundárias, a saber: radiação solar, declividade, curvatura, aspecto, índice topográfico combinado e amplitude altimétrica. A informação altitude, do MDE, foi inserida no modelo de classificação com o valor absoluto de cada célula do mapa gerado pela interpolação (Figura 6).

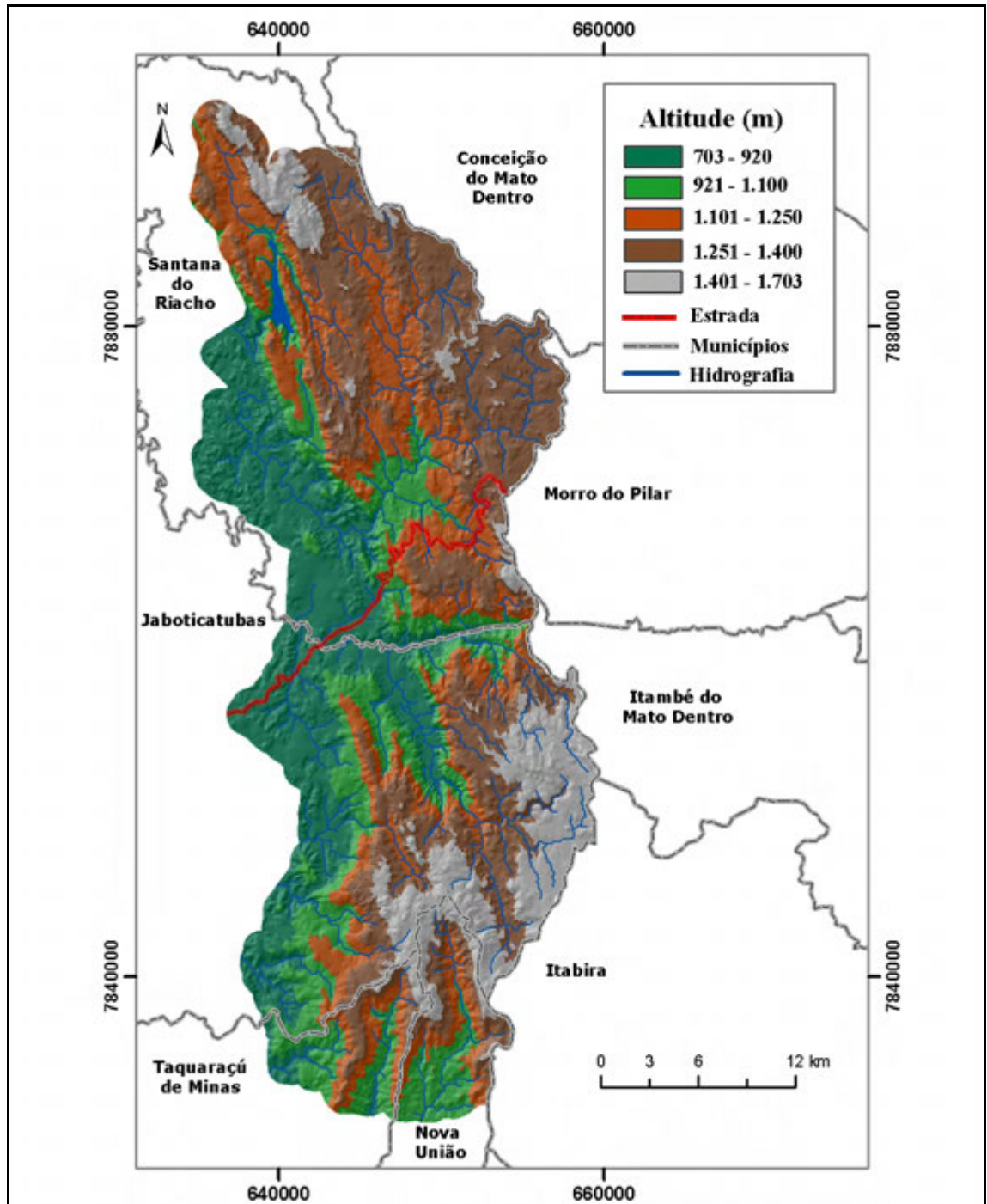


Figura 6 – Mapa de elevação do terreno.

3.4.2. Aspecto

Aspecto é definido como a direção da face da vertente, usualmente indicado pelo azimute ou direção cardinal, ou ainda como a direção da declividade. As áreas no mapa indicam a direção da taxa máxima de mudança no valor Z para cada célula central em relação às suas vizinhas, num plano de 3x3 células (CARVALHO JUNIOR, 2005). Os valores são expressos em graus positivos de 0 a 360, medidos a partir do norte, no sentido horário. O aspecto define a quantidade de radiação solar direta que uma área recebe, influenciando extremamente o microclima do solo. O mapa de aspecto foi derivado do MDE, utilizando-se o módulo *Spatial Analyst* do ArcGis 9.2 (Figura 7). No Quadro 2 são apresentadas as classes de aspecto da paisagem e a respectiva extensão de ocorrência na área em estudo.

Quadro 2 – Área das classes de aspecto

Classe	Área	
	(ha)	%
Plano	274,1	0,3
Norte	21.007,0	21,0
Leste	24.145,0	23,9
Sul	20.699,6	20,6
Oeste	34.588,6	34,3
Total	100.714,32	100

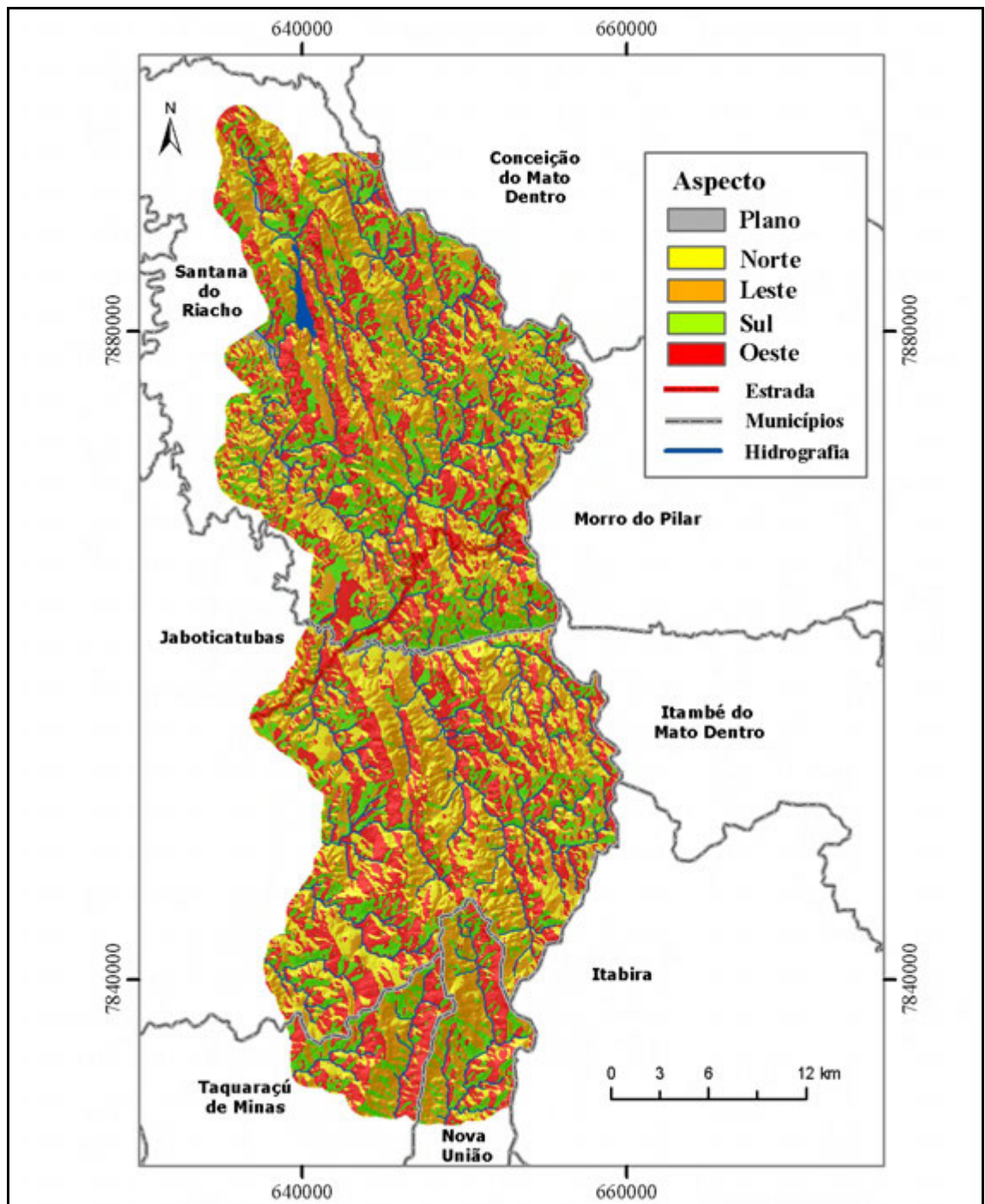


Figura 7 – Mapa de aspecto.

3.4.3. Declividade

O mapa de declividade foi gerado em porcentagem e reclassificado em seis classes para o mapeamento de solos, de acordo com os intervalos de declividade propostos pela Embrapa (1997), quais sejam: Plano (0-3), Suave Ondulado (3-8), Ondulado (8-20), Forte Ondulado (20-45), Montanhoso (45-75) e Escarpado (>75)

(Figura 8). A maior parte da área, 33%, possui declividade ondulada; 28%, forte ondulada; 20%, suave ondulada e, em menor proporção, ocorrem áreas planas, montanhosa e escarpadas (Quadro 3).

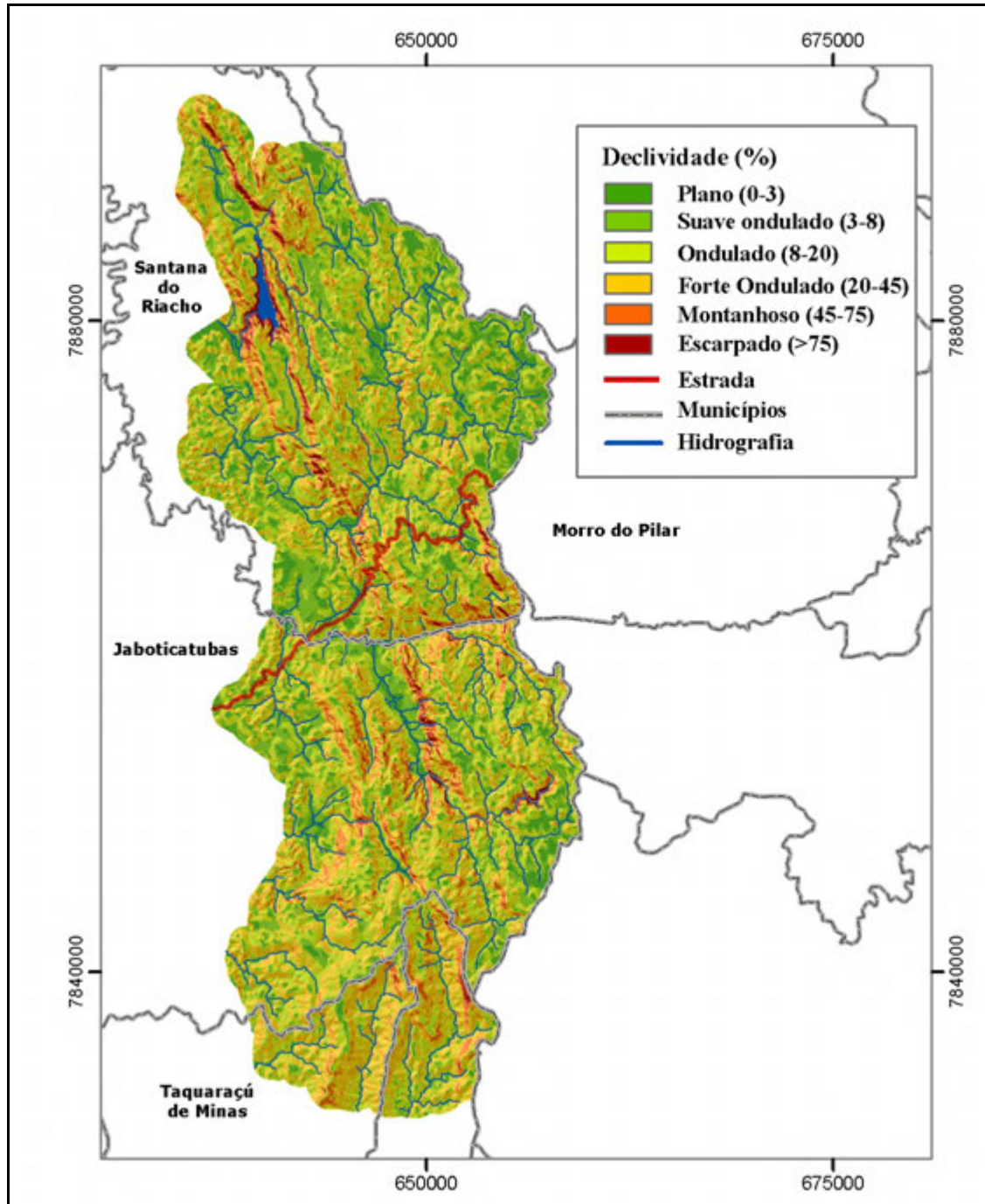


Figura 8 – Mapa de declividade.

Quadro 3 – Área das classes de declividade

Classe	Área	
	(ha)	(%)
Plano	12.805,6	12.7
Suave Ondulado	19.745,9	19.6
Ondulado	32.823,7	32.6
Forte Ondulado	27.754,5	27.5
Montanhoso	6.456,9	6.4
Escarpado	1.127,8	1.1
Total	100.714,3	100

3.4.4. Amplitude altimétrica

A topografia da área em estudo apresenta expressiva variação altimétrica, com os pontos mais altos nos divisores de duas importantes bacias hidrográficas: bacia do São Francisco, a leste, e bacia do Rio Doce, a oeste. Essa topografia, com forte influência na gênese dos solos, apresenta diferença bastante pronunciada entre a altitude das partes mais elevadas e aquelas dos vales. Considerando-se essas diferenças na paisagem, testou-se o uso da variável amplitude altimétrica para discriminar os solos em função da altitude.

Para tanto, foi utilizado o MDE para derivar os mapas de direção da drenagem e de fluxo acumulado do escoamento. A partir desses dois mapas foram geradas as bacias de contribuição de cada célula no mapa e, para essas, calculada a amplitude altimétrica. Em cada bacia, o valor de menor altitude, definido como a célula de maior fluxo acumulado da drenagem, foi subtraído dos valores de altitude das demais células. O procedimento foi feito utilizando-se álgebra de mapas, entre o modelo de elevação e o mapa de bacias. O mapa obtido é apresentado na figura 9.

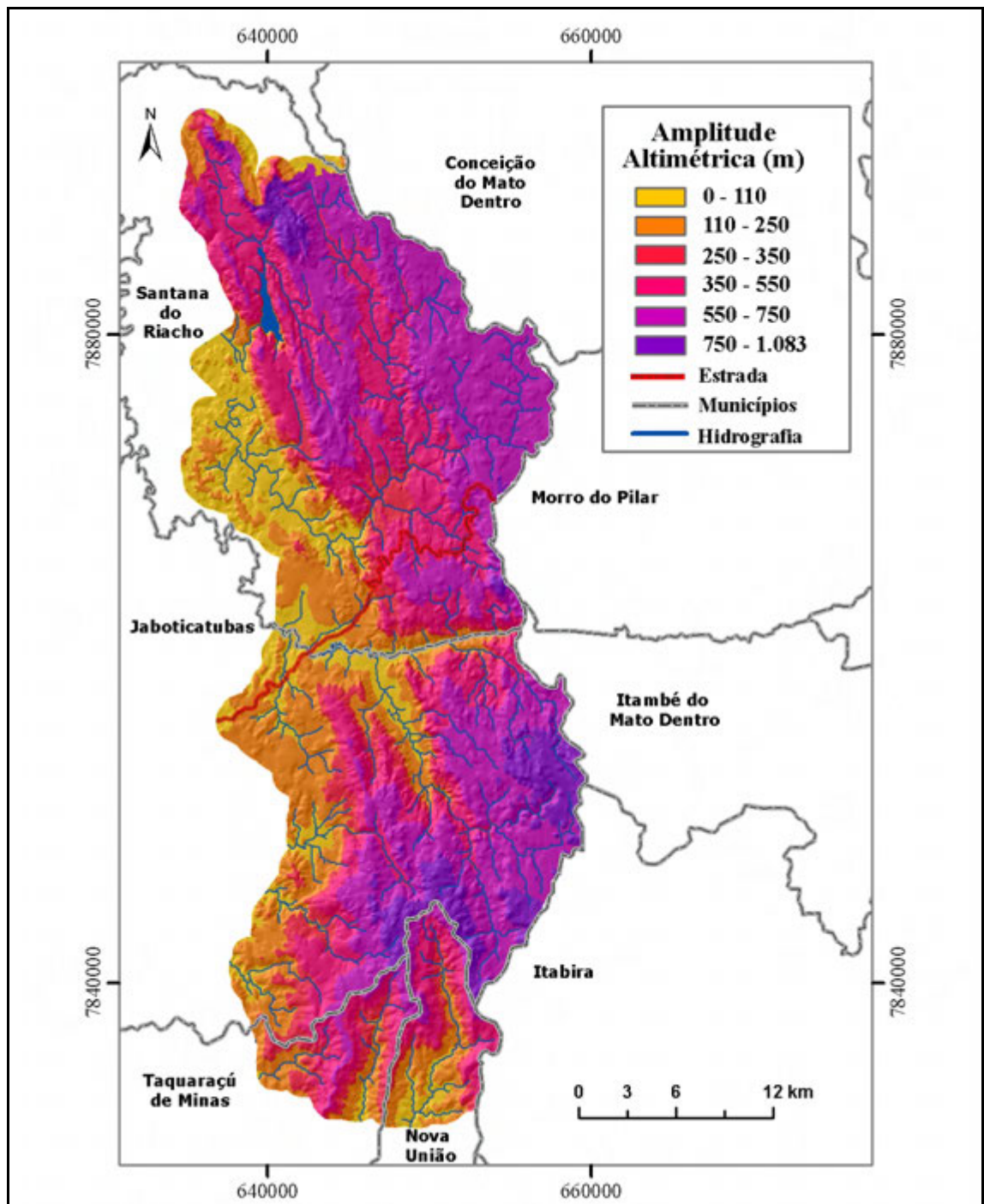


Figura 9 – Mapa de amplitude altimétrica.

3.4.5. Curvatura

O mapa de curvatura gerado foi reclassificado em duas classes. Os valores negativos até o valor zero foram considerados como áreas côncavas, e os valores positivos, como áreas convexas. No mapa em questão foram sobrepostas as áreas consideradas planas, as quais foram definidas como aquelas de relevo plano, ou seja,

com declividade entre 0 e 8 %, e, com distância de até 150 metros da hidrografia, para evitar que os topos de morro pudessem ser classificados nesta categoria. Assim, o mapa foi definido com três classes: plano, convexo e côncavo (Figura 10).

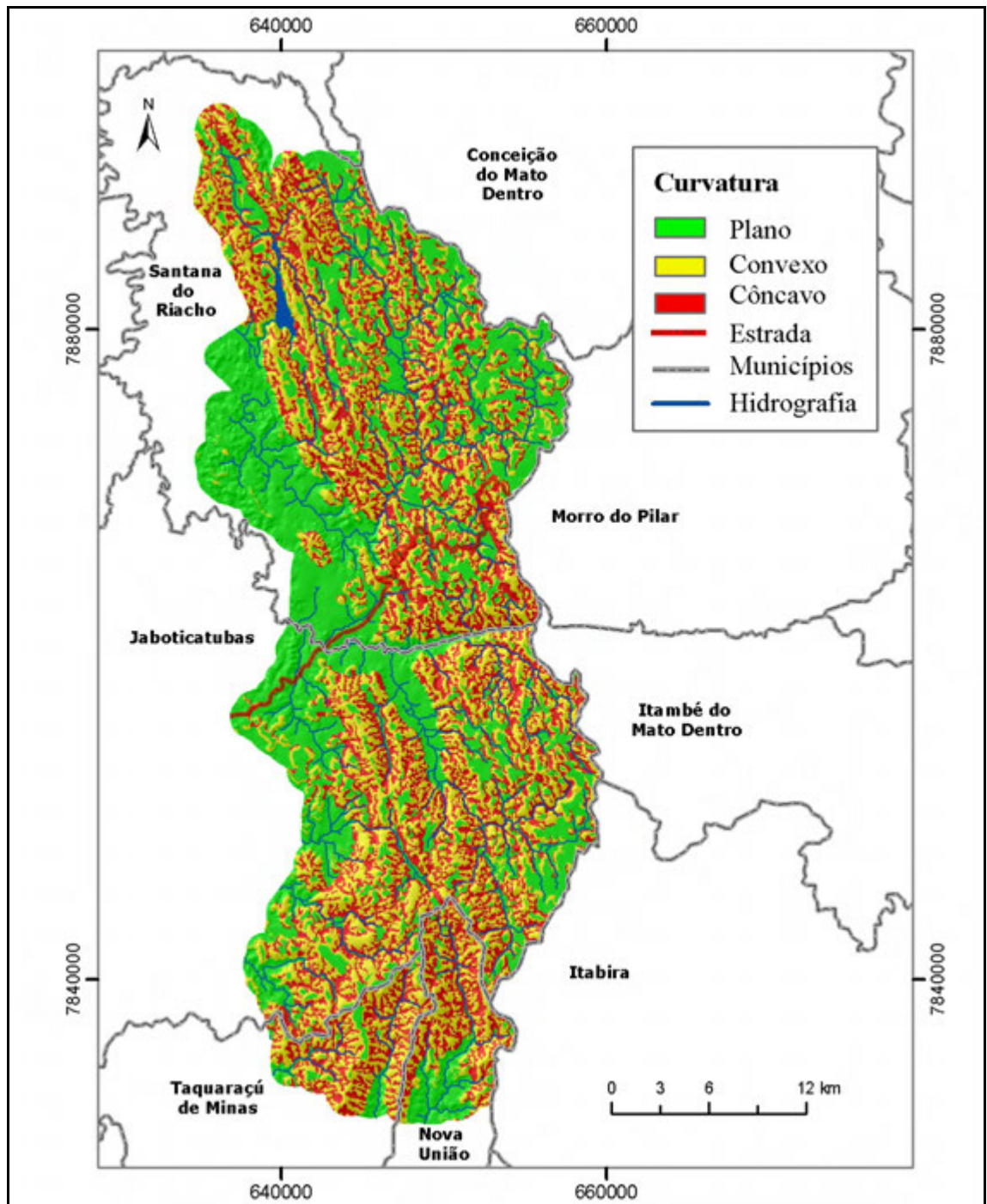


Figura 10 – Mapa de curvatura.

3.4.6. Radiação solar

A radiação solar influencia a gênese dos solos pelo fator clima e organismos. A intensidade desta está relacionada com a altitude e a morfologia da paisagem. A partir do MDE foi calculada a radiação global anual em um mapa onde cada célula do grid contém informação em Watt hora por metro quadrado (Wh/m^2). O mapa de radiação solar em Megawatt hora $/\text{m}^2/\text{ano}$ é apresentado na figura 11. Ele foi gerado utilizando-se o módulo *Solar analyst* do ArcGis 9.2.

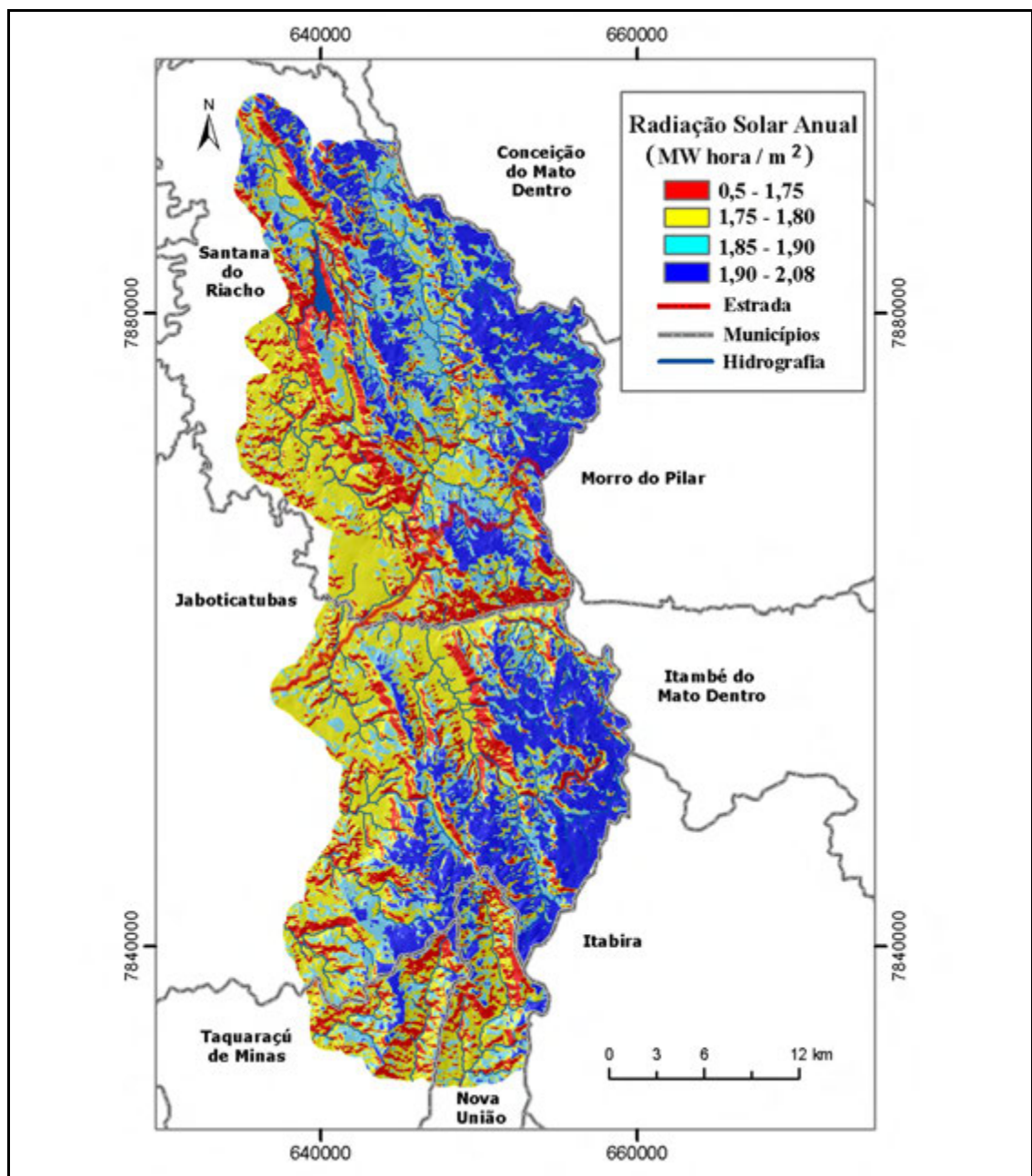


Figura 11 – Mapa de radiação solar.

3.4.7. Índice Topográfico Combinado

O CTI, gerado a partir das informações do MDE (Figura 12), é definido como uma função da declividade e da área de contribuição por unidade de largura ortogonal à direção do fluxo, segundo a equação 1:

$$CTI = \ln \frac{(As)}{(\tan\beta)} \quad (1)$$

Em que: (As) representa área de contribuição obtida pelo (fluxo acumulado + 1) multiplicado pelo tamanho da célula do grid, em m² e, (β) é o valor da declividade expressa em radianos.

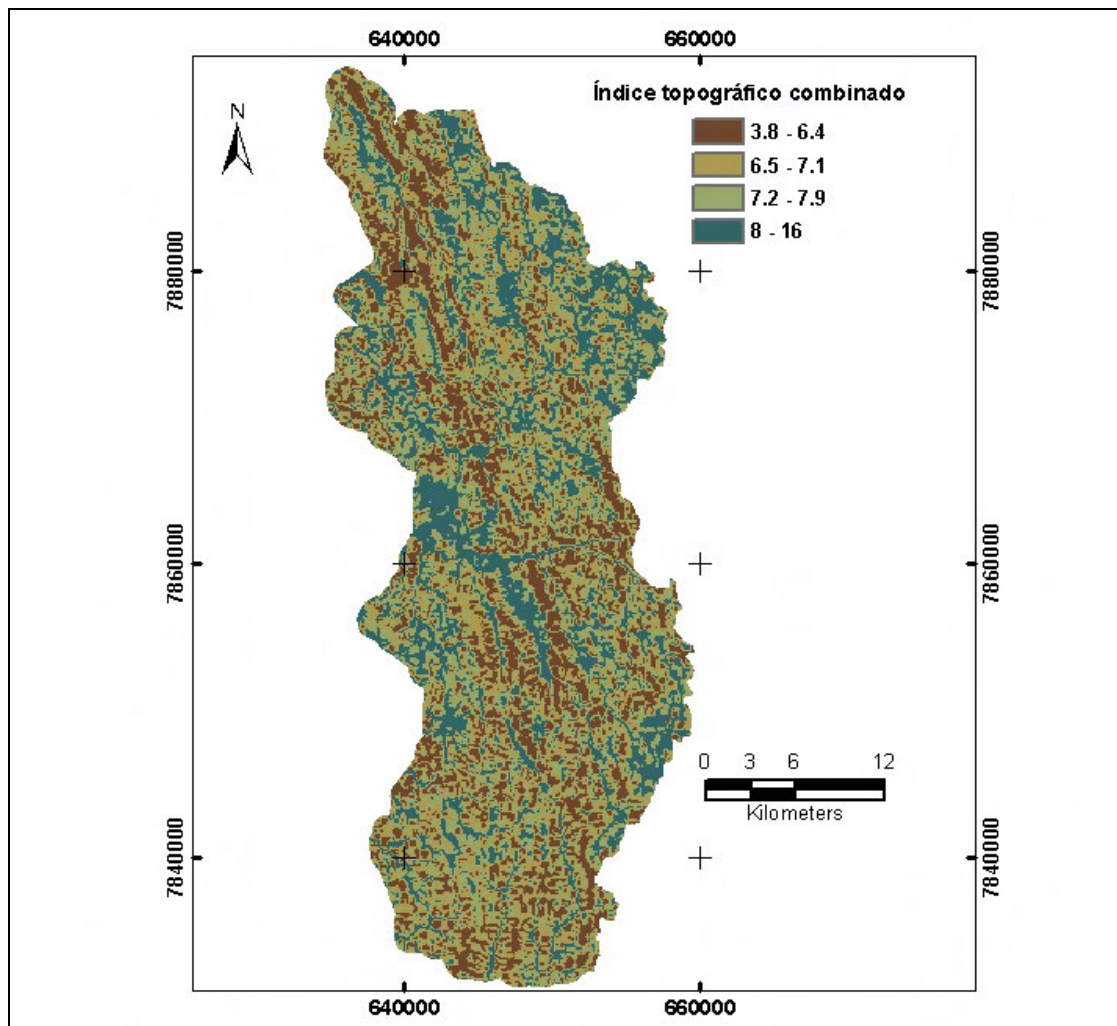


Figura 12 – Mapa de Índice topográfico combinado.

3.4.8. Geologia

A informação geológica utilizada foi extraída dos mapas geológicos em escala de 1:100.000 (COMIG, 1997), com recobrimento de 80% da área em escala de 1:1.000.000 (CPRM, 2004). Realizou-se a união dos dois mapas e, então, procedeu-se à análise e eliminação das incertezas.

Amostragens litológicas feitas durante os trabalhos de campo, integradas à interpretação de imagem de satélite e do MDE, possibilitaram a edição dos limites das unidades do mapa geológico e, assim, minimizar a propagação de erros durante a classificação de solos. O mapa obtido é apresentado na figura 13. No Quadro 4 estão as classes e respectivas áreas do mapa geológico.

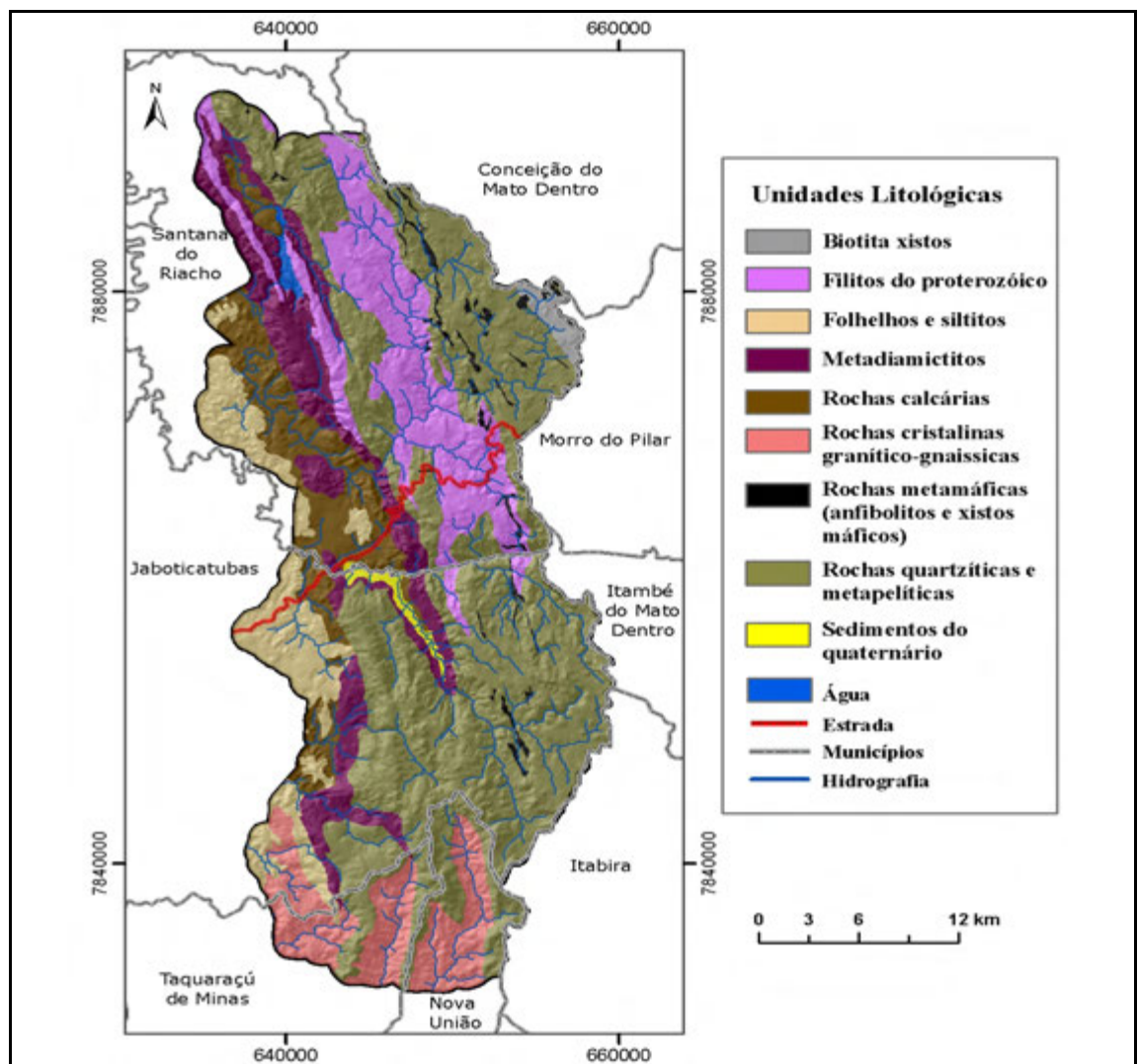


Figura 13 – Mapa de unidades litológicas da área em estudo, elaborado por COMIG (1997) e CPRM (2004).

Quadro 4 – Unidades litológica da área em estudo

Litologia	Área	
	(ha)	(%)
Rochas cristalinas granítico-gnaissicas	8.143	8,1
Filitos do proterozóico	13.983	13,9
Metadiamicititos	11.107	11,0
Rochas metamáficas - anfibolitos e xistos máficos	1.382	1,4
Rochas quartzíticas e metapelíticas subordinadas	46.809	46,5
Rochas calcárias	9.183	9,1
Biotita xistos	812	0,8
Sedimentos do quaternário - areias e cascalhos	705	0,7
Folhelhos e siltitos	8.247	8,2
Água	344	0,3
Total	100.714	100

3.4.9. Dados espectrais de imagem de satélite orbitais e índices derivados

O uso de dados de imagens espectrais é comumente feito tanto na forma primária de obtenção da imagem pelos valores do nível de cinza, que é a resposta das feições mapeadas, quanto na forma processada, por meio de índices derivados.

As informações espectrais de seis bandas do sensor ETM+ da faixa do visível, infravermelho próximo e médio foram empregadas com valores absolutos de nível de cinza (Quadro 5). Foram também utilizados como variáveis discriminantes quatro índices derivados dessas bandas espectrais, quais sejam: *Clay minerals*, *Ferrous minerals*, *Iron oxide* e índice de vegetação por diferença normalizada (NDVI).

Quadro 5 – Características das bandas espectrais da imagem do satélite Landsat, sensor ETM + utilizadas na classificação dos solos

Banda	Resolução espacial (m x m)	Faixa espectral (μm)	Região do espectro
1	30 x 30	0,450 – 0,515	Azul
2	30 x 30	0,525 – 0,605	Verde
3	30 x 30	0,630 – 0,690	Vermelho
4	30 x 30	0,750 – 0,900	Infravermelho próximo
5	30 x 30	1,550 – 1,750	Infravermelho médio
7	30 x 30	2,090 – 2,350	Infravermelho médio

O NDVI ressalta os solos em relação ao comportamento espectral da vegetação. O índice está baseado nas diferenças quantitativas entre a radiação refletida e transmitida por esses dois alvos (MOREIRA, 2003). A composição NDVI é feita por meio de uma operação matemática entre as bandas da região espectral do vermelho e a banda da região espectral do infravermelho próximo, conforme a equação 2:

$$\text{NDVI} = \frac{(\text{IVP} - \text{VER})}{(\text{IVP} + \text{VER})} \quad (2)$$

Em que, IVP = Banda espectral da faixa do infravermelho próximo e, VER = Banda espectral da faixa do vermelho.

A análise do NDVI mostra que, quanto maior, o vigor da vegetação, maior é o contraste entre as bandas usadas para a composição do índice.

A partir das bandas espectrais 1, 3, 4, 5 e 7, da imagem do satélite Landsat foram derivados índices para ressaltar o comportamento de atributos do solo. Esses índices são obtidos por operações matemáticas entre bandas espectrais como se seguem: *Clay minerals* (5/7), *Iron oxide* (3/1) e *Ferrous minerals* (5/4). Na figura 14 são apresentados os mapas de índices gerados.

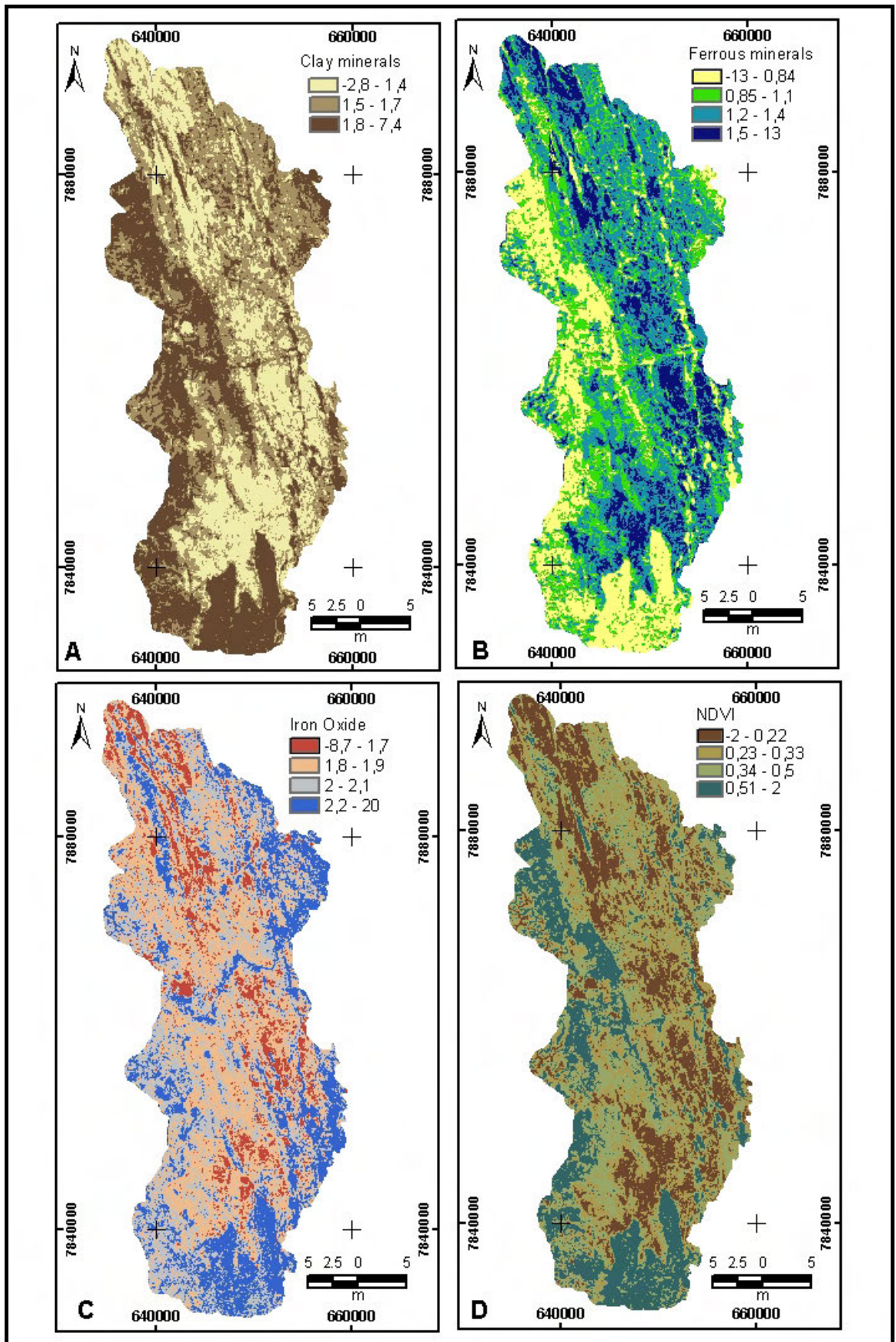


Figura 14 – Mapas de índices derivados das imagens ETM+: A) *Clay minerals*, B) *Ferrous minerals*, C) *Iron Oxide* e D) *NDVI*.

3.4.10. Informação pedológica

Dados pré-existentes de solo, na forma de perfis ou de mapas, têm sido integrados a dados auxiliares e utilizados como *inputs* em modelos preditivos de solos, em escalas intermediárias de mapeamento (VOLTZ & WEBSTER, 1990; SCULL et al, 2005; KROL, 2006; ZHU, 2000; CARRÉ & BOETTINGER, 2008).

Os mapas de solos integram um sistema de conhecimento sobre os solos, a sua distribuição espacial e o relacionamento destes com os fatores ambientais. O mapeamento digital busca extrair o conhecimento sistematizado, pelo pedólogo, em mapas e transferi-lo para modelos automatizados (BUI, 2004).

Em estudo realizado por Carré et al. (2008), a informação pedológica extraída do mapa convencional foi inserida no modelo de classificação supervisionada de classes de solo. A comparação entre o mapa obtido nesse estudo, com outro realizado sem o uso da informação pedológica, mostrou maior nível de detalhe da informação no mapa que utilizou informações previamente sistematizadas em mapa.

Zhu (2000) realizou mapeamento de solos utilizando informações extraídas de mapas de solos em conjunto com dados auxiliares. O resultado do mapeamento, empregando redes neurais, mostrou que o mapa gerado apresentava maior detalhamento da informação mapeada, comparando-se aquelas presente no mapa obtido por método convencional.

O mapa pedológico gerado pelo Projeto RADAM BRASIL foi utilizado como uma das variáveis discriminantes dos solos inserida no modelo de classificação por RNA como variável categórica. As informações foram extraídas da Folha Belo Horizonte (não publicado). O mapa apresenta informações pedológicas em escala de 1:1000.000, sendo composto por cinco unidades de solos classificados até o segundo nível categórico, e uma unidade representada pelos afloramentos de rochas (Figura 15).

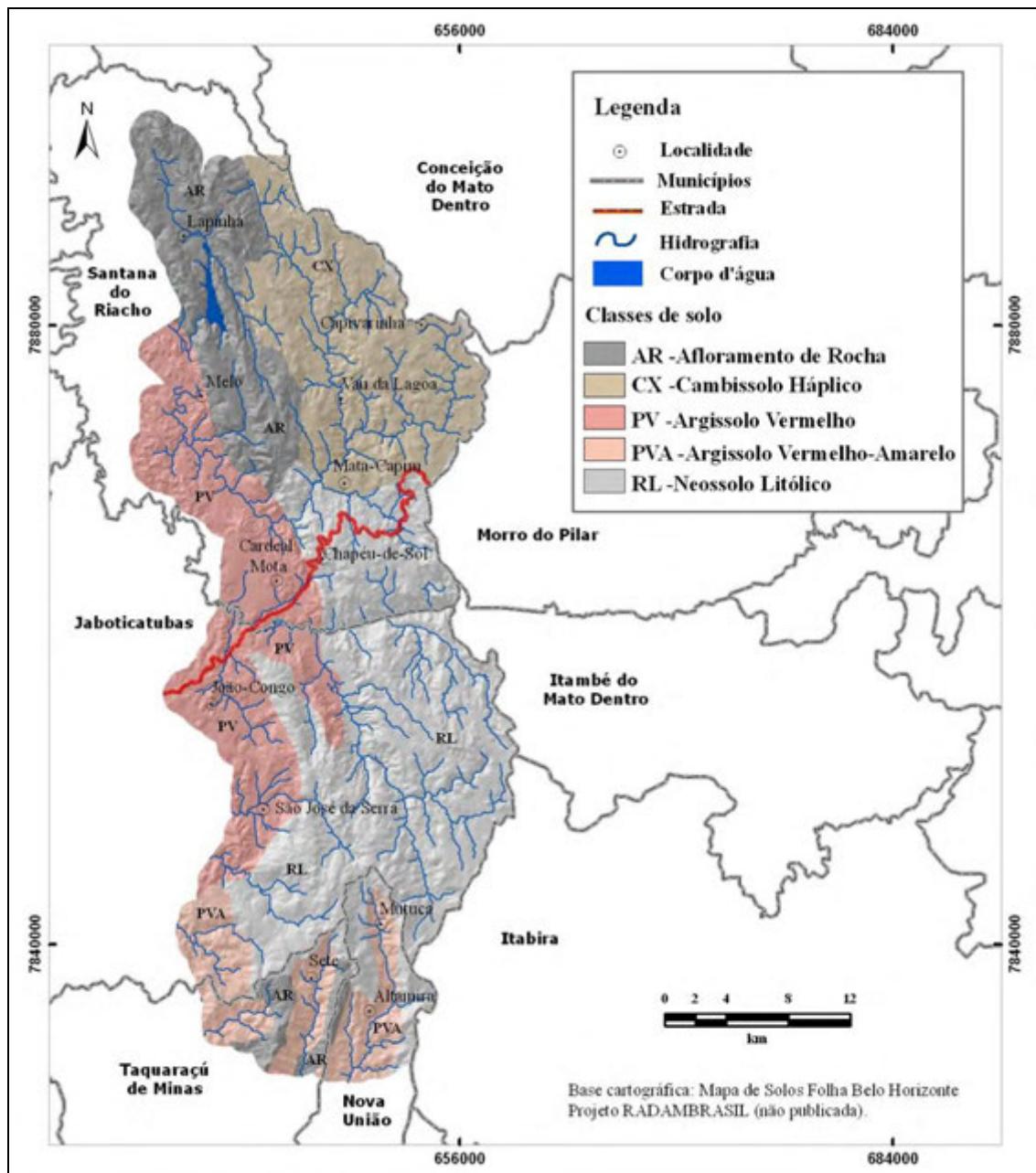


Figura 15 – Mapa de solos da área em estudo elaborado pelo Projeto RADAMBRASIL.

As unidades de mapeamento foram reclassificadas de acordo com o atual Sistema Brasileiro de Classificação de Solos (SiBCS, 2006) até o segundo nível categórico. Os solos associados às unidades de solos foram desconsiderados. No quadro 6 estão apresentadas a distribuição espacial das unidades, assim como a correspondente classificação dos solos no atual sistema.

Quadro 6 – Unidades de solos mapeadas na área em estudo pelo Projeto RADAM BRASIL na escala de 1:1.000.000

Unidade de mapeamento de solo	Classificação antiga	Área (ha)
AR - Afloramento de Rocha	AR – Afloramento de rocha	15.229
PV - Argissolo Vermelho	PEe - Podzólico Vermelho escuro álico	19.914
PVA - Argissolo Vermelho-Amarelo	PVa– Podzólico Vermelho álico	9.055
CX - Cambissolo Háptico	Ca – Cambissolo álico	18.663
RL - Neossolo Litólico	RLd, Ra – Solo Litólico	37.673
Total		100.715

3.5. Amostragem de padrões

As variáveis discriminantes selecionadas compuseram uma matriz para extração das amostras de treinamento e validação do classificador. A resolução espacial dessa matriz foi definida com célula de 30 m, tendo como base a resolução espacial das imagens de satélite e a equidistância das curvas de nível.

A amostragem para treinar e validar a classificação foi realizada tendo como orientação a distribuição dos perfis de solo na paisagem. Para tanto, foi gerada uma imagem de referência, na qual cada unidade de solo teve sua área de ocorrência delimitada. Para cada unidade, amostras de todas as variáveis discriminantes foram coletadas dentro do limite das respectivas áreas na imagem de referência. A coleta foi feita por região, de forma aleatória.

O tamanho da amostra para classificação e validação do treinamento foi definido segundo critérios estatísticos, sendo dimensionado entre dez a 30 vezes o número de variáveis discriminantes (FOODY & ARORA, 1997).

O classificador de redes neurais é capaz de generalizar satisfatoriamente padrões com poucas amostras, desde que estas representem adequadamente os padrões de treinamento; já o MaxVer necessita de um tamanho mínimo de amostra que permita estatisticamente discriminar as variáveis por meio de dados estatísticos (FOODY & ARORA, 1997). Assim, a amostragem visou à classificação pelos dois classificadores com o mesmo conjunto de dados, possibilitando a comparação entre os mapas gerados pelo índice de exatidão kappa, conforme sugere Mather (2004).

3.6. Treinamento do classificador e definição da arquitetura das redes neurais

A classificação pelas RNAs foi realizada utilizando-se o simulador *Stuttgart Neural Network Simulator* (ZELL et al., 2008). Foi empregado o modelo de redes multicamadas e o algoritmo de aprendizagem de retropropagação (HIROSE et al., 1991).

A arquitetura das redes neurais foi construída seguindo uma estrutura conceitual em que a camada de entrada foi conectada à camada de saída por uma ou mais camadas internas. Nessa arquitetura, o número de neurônios da camada de entrada correspondeu ao número de variáveis discriminantes utilizadas, enquanto que, na camada de saída, o número de neurônios foi definido pelo total de classes de solo. O número de camadas internas, bem como o de neurônios destas camadas, foram determinados por tentativa e erro baseando-se no método construtivo (HIROSE et al., 1991).

Os valores das variáveis foram escalonados para intervalo entre 0 e 1. Esse procedimento foi necessário devido à grande variação nas unidades de medida das variáveis, e visa facilitar os processamentos da classificação durante os ajustamentos dos pesos atribuídos no treinamento, que ocorrem de forma interativa. Assim, evita-se que ocorra a saturação das redes durante a propagação dos valores atribuídos nos neurônios de processamentos das redes (KANNELLOPOULOS & WILKINSON, 1997).

Procedimento adicional, constou da conversão dos arquivos com as informações dos solos, extraídas dos mapas de variáveis discriminantes, para arquivos utilizados no programa de redes neurais. Utilizaram-se os aplicativos desenvolvidos por Vieira (2000). O treinamento do classificador envolveu testes para determinar o número de unidades de solos, a escolha da arquitetura (neurônios e camadas), os parâmetros da rede e o desempenho das variáveis na discriminação dos solos.

O algoritmo de retropropagação foi treinado com a taxa de aprendizado variando de 0,2 a 0,1 ($\eta = 0,2$ e $d_{max} = 0,1$). Os pesos foram iniciados com valor mínimo de -0,5 e máximo de 0,5, sendo recalculados pela função de ordem aleatória. O número de ciclos de treinamento foi avaliado de acordo com o quadrado médio do erro (MSE).

As unidades de solo para classificação foram estabelecidas em função da escala de mapeamento dos solos definida como sendo 1:100.000, com base no detalhe cartográfico da base de dados e o número de perfis amostrado. As unidades de solo foram compostas por classes simples ou por associação de classes de solos, sendo as

unidades definidas durante o treinamento da classificação, de acordo com o grau de dificuldade do classificador em mapear os solos.

As variáveis discriminantes dos solos foram selecionadas para a classificação por meio de testes estatísticos observando-se a contribuição de cada uma de forma individual e por agrupamento. Na análise individual, as variáveis foram excluídas uma por vez; já na análise de grupos, variáveis similares (dados espectrais e atributos do terreno) foram agrupadas e excluídas do modelo em conjunto.

As classificações realizadas com os diferentes grupos de variáveis pelos dois classificadores foram submetidas ao Teste Z para observar a diferença estatística significativa entre as variáveis e classificadores, conforme Congalton e Green (1999).

Após o treinamento do classificador com as amostras informacionais para cada uma das unidades de solo, a definição dos parâmetros e a arquitetura das redes neurais a partir de avaliação dos resultados, foi realizada a classificação propriamente dita, ou seja, a generalização da informação treinada com as amostras para toda a imagem.

O treinamento do classificador foi iniciado com a arquitetura das redes neurais definida por uma camada de entrada composta pelas 19 variáveis discriminantes, e a camada de saída com 13 unidades de solo. Algumas unidades de solo foram associadas durante o processo de treinamento das redes neurais, devido à dificuldade do classificador em separá-las.

3.7. Validação dos mapas

A validação dos mapas classificados pelas redes neurais, MaxVer e mapas obtidos pelo método convencional foi feita utilizando-se 362 pontos de solos identificados e georreferenciados para essa finalidade (Figura 16). Os Espodossolos e Organossolos identificados durante o levantamento, em áreas de platôs soerguidos, com o auxílio de helicóptero, não tiveram pontos coletados para validação dos mapas classificados, devido à impossibilidade de acesso por via terrestre e à inviabilidade de uso de outro meio de transporte para visitar a área.

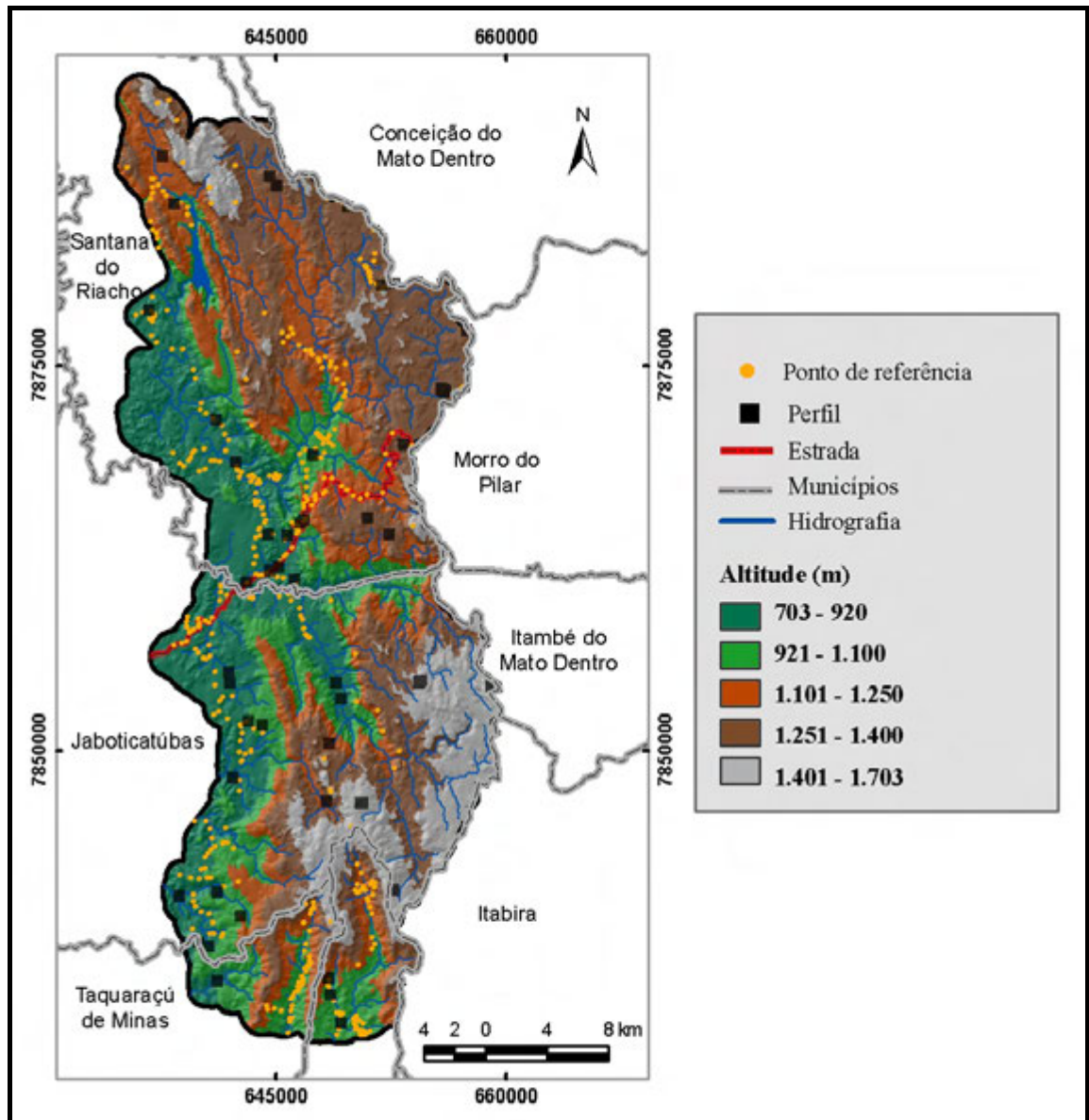


Figura 16 – Localização dos perfis de solos e pontos de referência utilizados na classificação e validação do mapa de solo classificação.

A validação da classificação foi feita utilizando-se a técnica multivariada por meio da matriz de confusão, a qual tem como indicadores de exatidão o índice kappa e a exatidão global.

A exatidão global (P_o) representa a concordância geral da classificação. O seu cálculo é feito dividindo-se o total de unidades amostrais classificadas corretamente pelo total das amostras (Equação 3).

$$P_o = \frac{\sum_{i=1}^m n_{ii}}{N} \quad (3)$$

Em que: (N) representa o número total de unidades amostrais contempladas pela matriz; (n_{ii}) representa os elementos da diagonal principal e (m) simboliza o número de classes da matriz.

A partir dos resultados presentes na matriz de confusão pode-se derivar os erros de omissão e comissão. O erro de omissão é uma medida de incerteza que representa a exatidão do produtor e refere-se à probabilidade da amostra está corretamente classificada. Ele é calculado dividindo-se o total de pixels corretamente classificados numa classe, pelo número total de pixels da classe. No cálculo da exatidão de uma determinada classe, serão omitidas do cálculo as amostras assinaladas como pertencentes à classe de análise, mas que foram classificadas em outras classes.

Por outro lado, a exatidão do usuário representa o erro de comissão, ou seja, a probabilidade de um pixel classificado como pertencente a uma classe no mapa realmente pertencer a tal classe na área de mapeamento. O cálculo do erro de comissão é feito a partir da divisão do total de amostras corretamente classificado na classe pelo total de amostra assinalada como pertencente à classe.

O índice kappa, que também pode ser derivado da matriz de confusão, diferentemente da exatidão global, utiliza todos os dados da matriz para o cálculo da concordância do mapeamento. Este índice é indicado por Congalton e Green (1999) como o indicador mais adequado para a avaliação da exatidão de classificação, uma vez que ele não considera apenas a exatidão global, mas os erros de comissão e omissão, permitindo a estimativa da exatidão posicional.

O índice kappa varia de (-1) a (1). Os valores próximos de zero indicam baixa concordância entre o mapa gerado e as informações de campo, enquanto os próximos de 1 indicam a maior concordância da classificação.

O cálculo do índice kappa é feito conforme a equação 4, em que o (P_o) é a exatidão global; (P_e) é calculado segundo a equação 5. Na equação 5, (n_{i+}) e (n_{+i}) representam as proporções marginais da linha (i) e da coluna (i), respectivamente.

$$K = \frac{P_o - P_e}{1 - P_e} \quad (4)$$

$$P_e = \frac{\sum_{i=1}^M n_{i+} \times n_{+i}}{N^2} \quad (5)$$

Assim como calculados para o mapa, o índice kappa (Kappa condicional K_i) e a variância podem ser calculados para cada uma das classes da matriz conforme as equações 6 e 7, respectivamente, para o kappa e a variância. Os valores desses índices, obtidos por diferentes abordagens de classificação, permitem avaliar a existência de diferença estatística significativa na exatidão de classificações feitas por diferente métodos ou conjunto de variáveis. Tendo sido usado nas comparações feitas pelas duas abordagens testadas, o MaxVer e as Redes Neurais.

$$\hat{K}_i = \frac{nn_{ii} - n_{i+}n_{+i}}{nn_{i+} - n_{i+}n_{+i}} \quad (6)$$

$$V\hat{a}r(\hat{K}) = \frac{n(n_{i+} - n_{ii})}{[n_{i+}(n - n_{+i})]^3} [(n_{i+} - n_{ii})(n_{i+}n_{+i} - nn_{ii}) + nn_{ii}(n - n_{i+} - n_{+i} + n_{ii})] \quad (7)$$

O índice kappa pode ser analisado segundo os seguintes intervalos: ($K > 0,4$), baixa concordância, ($0,4 > K > 0,8$), concordância média e, ($K > 0,8$) alta concordância (LANDIS & KOCH, citados por CONGALTON e GREEN, 1999).

A concordância entre o mapa classificado pelo método convencional (Figura 17) e os pontos de referência terrestre foi obtida utilizando álgebra de mapas, conforme proposto por Boots e Csillag (2006). A exatidão global (P_o) foi o índice empregado para a análise da concordância. No quadro 7 estão apresentadas as unidades de solo do e as suas respectivas áreas de ocorrência.

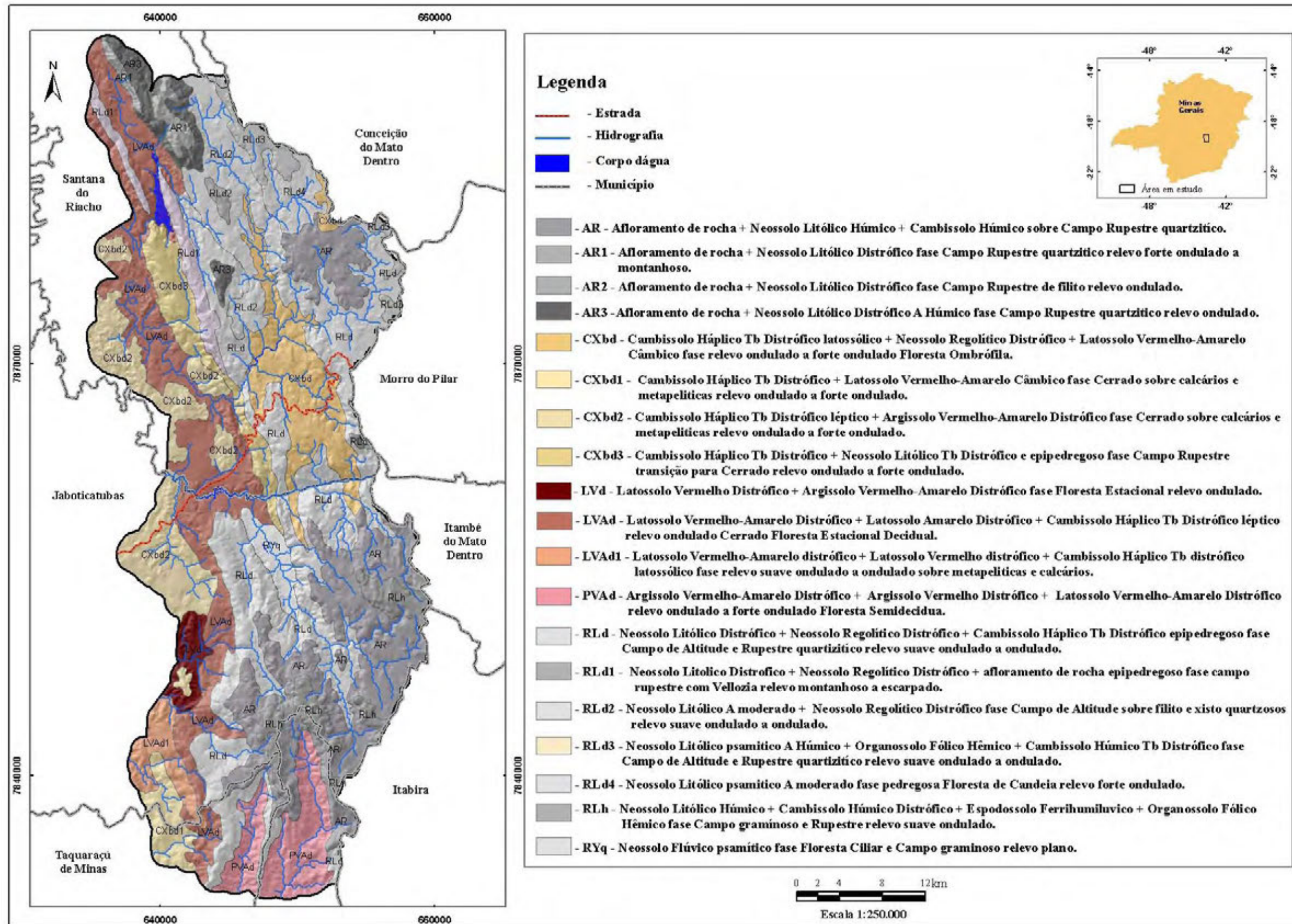


Figura 17 – Mapa de solos obtido pelo Método convencional (Schaefer et al., 2008)

Quadro 7 – Unidades de solo mapeadas pelo método convencional

Unidade de Mapeamento de solo	Área (ha)
AR - Afloramento de rocha + Neossolo Litólico Húmico + Cambissolo Húmico sobre Campo Rupestre quartzítico.	15.574
AR1 - Afloramento de rocha + Neossolo Litólico Distrófico fase Campo Rupestre quartzítico relevo forte ondulado a montanhoso.	1.096
AR2 - Afloramento de rocha + Neossolo Litólico Distrófico fase Campo Rupestre de filito relevo ondulado.	296
AR3 - Afloramento de rocha + Neossolo Litólico Distrófico A Húmico Campo Rupestre quartzítico relevo ondulado.	2.259
CXbd - Cambissolo Háplico Tb Distrófico latossólico + Neossolo Regolítico Distrófico + Latossolo Vermelho-Amarelo Câmbico fase relevo ondulado a forte ondulado Floresta Ombrófila.	7.225
CXbd1 - Cambissolo Háplico Tb Distrófico + Latossolo Vermelho-Amarelo Câmbico fase Cerrado sobre calcários e metapelíticas relevo ondulado a forte ondulado.	2.038
CXbd2 - Cambissolo Háplico Tb Distrófico léptico + Argissolo Vermelho-Amarelo Distrófico fase Cerrado sobre calcários e metapelíticas relevo ondulado a forte ondulado	7.679
CXbd3 - Cambissolo Háplico Tb Distrófico + Neossolo Litólico Tb Distrófico e epipedregoso fase Campo Rupestre transição para Cerrado relevo ondulado a forte ondulado.	3.411
LVAd - Latossolo Vermelho-Amarelo Distrófico + Latossolo Vermelho Distrófico + Cambissolo Háplico Tb Distrófico léptico relevo ondulado Cerrado Floresta Estacional Decidual.	15.098
LVA1 - Latossolo Vermelho-Amarelo distrófico + Latossolo Amarelo distrófico + Cambissolo Háplico Tb distrófico latossólico fase relevo suave ondulado a ondulado sobre metapelíticas e calcários.	1.962
LVd - Latossolo Vermelho Distrófico + Argissolo Vermelho-Amarelo Distrófico fase Mata Estacional relevo ondulado.	1.328
PVAd - Argissolo Vermelho-Amarelo Distrófico + Argissolo Vermelho Distrófico + Latossolo Vermelho-Amarelo Distrófico relevo ondulado a forte ondulado Floresta Semidecidua.	5.090
RLd - Neossolo Litólico Distrófico + Neossolo Regolítico Distrófico + Cambissolo Háplico Tb Distrófico epipedregoso fase Campo de Altitude e Rupestre quartzítico relevo suave ondulado a ondulado.	21.415
RLd1 - Neossolo Litólico Distrófico + Neossolo Regolítico Distrófico + afloramento de rocha epipedregoso fase campo rupestre com Vellozia relevo montanhoso a escarpado.	2.121
RLd2 - Neossolo Litólico A moderado + Neossolo Regolítico Distrófico fase Campo de Altitude sobre filito e xisto relevo suave ondulado a ondulado.	5.478
RLd3 - Neossolo Litólico psamítico A Húmico + Organossolo Fólico Hêmico + Cambissolo Húmico Tb Distrófico fase Campo de Altitude e Rupestre quartzítico relevo suave ondulado a ondulado.	1.771
RLd4 - Neossolo Litólico psamítico A moderado fase pedregosa Mata de Candeia relevo forte ondulado.	1.734
RLh - Neossolo Litólico Húmico + Cambissolo Húmico Distrófico + Espodossolo Ferrihumiluvico + Organossolo Fólico Hêmico fase Campo gramínoso e Rupestre relevo suave ondulado.	3.471
RYq - Neossolo Flúvico psamítico fase Floresta Ciliar e Campo gramínoso relevo plano.	1.389
Total	100.716

3.8. Comparação entre os mapas obtidos por diferentes classificadores

A concordância entre o mapa obtido na classificação pelas redes neurais e pelo MaxVer foi comparada empregando-se o teste Z. Esse teste utiliza valores do índice kappa e da variância para inferir a significância estatística da classificação em relação a uma classificação aleatória e na comparação entre duas ou mais classificações. Para tanto, foram obtidos os valores de Z e da variância para cada uma das classificações obtidas.

O cálculo da variância é feito conforme a equação 6, em que o (n) é o número de amostras e θ_1 a θ_4 calculado segundo as equações 7 a 10, onde k representa o número de classes da matriz. Na equação 7, (n_{ii}) representa os elementos da diagonal principal, (n_{i+} e n_{+i}), nas equações 8 e 9 representam as proporções marginais da linha (i) e da coluna (i), respectivamente. Na equação 10, (n_{ij}) refere-se ao elemento da matriz na posição da linha (i) e coluna (j) e (n_{j+}) e (n_{+i}) o total de amostras da coluna e linha marginais, respectivamente.

$$V\hat{ar} \left(\hat{K} \right) = \frac{1}{n} \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4} \right\} \quad (6)$$

$$\theta_1 = \frac{1}{n} \sum_{i=1}^k n_{ii} \quad (7)$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^k n_{i+}n_{+i} \quad (8)$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^k n_{ii} (n_{i+}n_{+i}) \quad (9)$$

$$\theta_4 = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^k n_{ij} (n_{j+} + n_{+i})^2 \quad (10)$$

O teste Z permite verificar se o resultado da classificação é significativo em relação a uma classificação feita de forma aleatória. Para esse teste foi adotado o nível

de significância de 0,05 (valor crítico é 1,96). Na interpretação, valores maiores do que esse limite são significativos, ou seja, a classificação apresenta resultados melhores do que os obtidos por uma classificação processada de forma aleatória. O valor de Z pode ser obtido conforme a equação 11. O valor do teste comparação entre duas classificações para avaliar a diferença estatística significativa entre elas é obtido conforme a equação 12. Os resultados indicam a existência ou não de significância entre a exatidão das classificações feitas por dois mais classificadores, conjunto de dados ou outras procedimentos que se queria avaliar. O teste foi utilizado para comparar os resultados obtidos para os mapas, de forma geral, e para as unidades dos mesmos, comparando-se as classificações feitas pelas redes neurais e pelo MaxVer.

$$Z = \frac{\hat{K}}{\sqrt{\text{vâr } \hat{K}}} \quad (11)$$

$$Z = \frac{\left| \hat{K}_1 - \hat{K}_2 \right|}{\sqrt{\text{vâr } \left(\hat{K}_1 \right) + \text{vâr } \left(\hat{K}_2 \right)}} \quad (12)$$

4. RESULTADOS E DISCUSSÃO

A classificação foi feita, inicialmente, considerando as 20 classes de solos mapeadas, como uma unidade. A partir dos resultados da validação com amostras de treinamento foram observados os confundimentos entre as classes e foram redefinidas novas unidades que, em seguida, foram reclassificadas. Esse processo foi realizado sucessivamente até que os treinamentos foram considerados satisfatórios, ao fim dos quais os solos foram agrupados em nove unidades, sendo três simples, constituídas por uma classe cada uma, e seis unidades compostas por duas ou mais classes (Quadro 8).

Quadro 8 – Descrição das unidades de solo

Unidade	Classe 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; Floresta Semidecidual, relevo ondulado/forte ondulado.	Calcário.
CXbd	Cambissolo Háptico Tb latossólico, textura média e argilosa, A moderado; Cambissolo HápticoTb 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 Háptico 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.

As unidades de solo definidas para a classificação pelas redes neurais foram as mesmas utilizadas para a classificação pelo MaxVer, visando realizar a comparação entre os mapas gerados pelos dois classificadores de forma direta, ou seja, sem considerar as diferenças no número de classe. O tamanho da amostra foi definido conforme Foody e Arora (1997), sendo um total de 400 amostras para o treinamento e 200 para a validação, para cada uma das unidades de solo. Essas amostras de validação foram utilizadas para gerar matrizes de confusão que possibilitaram avaliar o desempenho dos classificadores e definir os parâmetros finais da classificação, quais sejam: pelas redes neurais, o número de neurônios e ciclos de treinamento, e pelos dois classificadores, redes neurais e MaxVer, as variáveis que contribuíram na classificação.

As unidades de solo, conforme apresentado no quadro 8, compreendem solos de quatro Ordens do Sistema Brasileiro de Classificação, sendo três dessas, Argissolos, Latossolos e Neossolos, subdivididos em duas Subordens e, os Cambissolos, representados por quatro Subgrupos. Esses solos estão distribuídos em seis classes de geologia, com o relevo variando entre plano, ondulado a montanhoso e escarpado. A vegetação é outro fator de estratificação indireta desses solos, a qual se apresenta como resposta às condições do ambiente dada pelo material de origem dos solos e do clima.

A caracterização das unidades de solo, por cinco atributos do terreno avaliados na classificação, a partir dos valores estatísticos das amostras selecionadas para validação da classificação é apresentada no quadro 9. Em relação à altitude, os Latossolos associados a Argissolos apresentaram a maior variação, com solos ocorrendo entre 724 e 1.258 m de altitude. Os Neossolos Flúvicos ocorrem nas menores cotas, com aproximadamente 100 m de variação; enquanto os Neossolos Litólicos associados a afloramentos de rochas ocorrem nas posições mais elevadas da paisagem, em áreas acima de 1.000 m. Os Cambissolos Latossólicos (CXbd1), por sua vez, ocorrem em áreas de Floresta Nebular com altitude entre 1.300 e 1.500 m.

Quanto a declividade da área, a maior variação ocorreu na unidade CXbd, com valor máximo e mínimo respectivamente igual a 73 e 7%. Os Cambissolos, assim como os Afloramentos de rochas associados a Neossolos, apresentaram os valores mais altos de declividade e radiação solar. A variável amplitude altimétrica apresentou valores distintos entre as unidades, com acentuada variação intraclasses. Já o índice CTI, com mínimo de 4,8 e máximo de 13,9, apresentou valores mais próximos.

Quadro 9 – Valores dos atributos do terreno e das amostras de validação para as unidades de solo

Unidade	Altitude (m)			Declividade (%)			Amplitude altimétrica (m)			Radiação (MWm ²)		CTI	
	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Max	Min	Max
PVAd	856	933	1.013	6	12	29	47	112	216	1,77	1,90	6,1	9,3
PVd + PVAd	731	875	954	9	26	41	35	174	246	1,60	1,89	5,1	7,8
CXbd1	808	854	909	4	17	26	103	149	197	1,57	1,89	5,5	8,2
CXbd	1.325	1.428	1.504	7	22	73	627	722	794	1,56	2,02	4,8	8,1
LVd + LVA	993	1.111	1.258	12	28	36	143	266	429	1,62	1,86	5,6	7,7
LVd + PVd	724	844	1.107	1	5	11	25	145	408	1,76	1,88	6,7	13,9
RLd + RRd	1.093	1.233	1.348	5	16	30	396	532	650	1,58	1,96	5,5	9,4
RYd	738	781	831	1	3	7	40	71	120	1,77	1,82	7,3	12,7
AR+ RLd	1.399	1.558	1.660	4	17	42	280	777	977	1,81	2,05	5,2	8,4

Valores: Min (Mínimo), Max (Máximo) e Med (Médio).

No quadro 10 são apresentados para cada uma das unidades de solo os valores mínimo, médio e máximo dos índices derivados da imagem, quais sejam: *Clay minerals*, *Ferrous minerals* e *Iron oxide* e NDVI. Os valores correspondem àqueles das amostras de validação. A distribuição dos valores máximo desses índices é apresentada na figura 19. Observa-se que o NDVI apresentou a menor variabilidade de valores entre os solos. As unidades CXbd1 e AR + RLd apresentaram os menores valores, e menores variações de valores, com médias de 0,17 e 0,14 respectivamente para a primeira e segunda unidades. Por outro lado, a unidade CXbd, devido a ocorrência sob Floresta Nebular apresentou o valor mais alto para o índice (0,69). As unidades de Latossolos e Argissolos apresentaram valores similares entre si, variando de 0,2 a 0,67. Os índices *Ferro mineral* e *Iron oxide* mostraram a mesma tendência na distribuição dos valores.

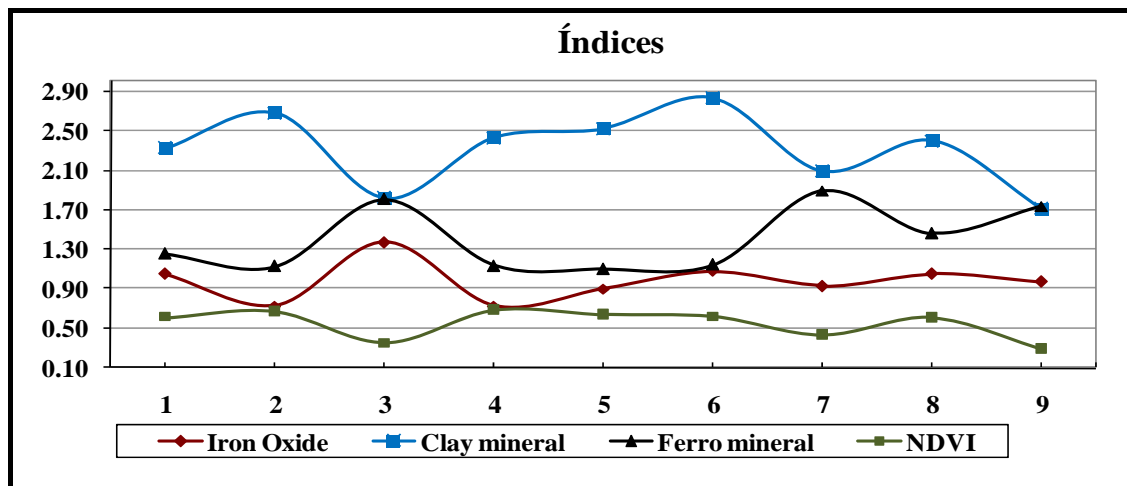


Figura 19 – Valor máximo dos índices derivados da imagem para as unidades de solo: 1))PVAd, 2)PVd + PVAd, 3)CXbd1, 4)CXbd, 5)LVd + PVd, 6)LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.

No quadro 11 são apresentadas as correlações entre atributos do terreno com valores contínuos e índices, avaliados na classificação das unidades de solo. As correlações foram calculadas com 1800 amostras selecionadas para a validação da classificação, sendo 200 amostras por unidade de solos. As oito variáveis selecionadas foram avaliadas com aquelas que se esperavam maior correlação entre unidades.

O índice *Clay mineral*, conforme observado no gráfico e na estatística dos valores apresentados no quadro 10 e nos coeficiente de correlação (Quadro 11) apresentou alta correlação positiva com o índice NDVI. A significância dessa correlação é alta nas unidades de Latossolos e Argissolos – com 99% de probabilidade – com mesmo nível de significância, obtendo R^2 de 0,92 com as amostras total das classes de validação.

A relação entre as variáveis altitude e amplitude altimétrica nas unidades de solo, conforme apresentadas no gráfico de valores máximo da variável na figura 19 e nos coeficientes de correlação linear (Quadro 12), não mostra que de forma geral a associação entre essas variáveis seja positiva e altamente significativa. Esse padrão não é observado para três unidades: os Argissolos de gnaisses e os Latossolos associados a Argissolo, de gnaisses e calcários, com valores não significativos, e os Neossolos Fúvicos, que apresentaram alta correlação significativa e negativa.

A correlação entre a declividade e a radiação solar global é inversa para a maioria das unidades, com alta significância. O coeficiente obtido para todas as amostras (-0.31) com significância maior do que 0,001 mostra essa tendência, com

quatro das unidades com coeficiente maior que 0,60. A exceção ocorreu para os Cambissolos (CXbd) com R^2 não significativo a 10%, e para os Latossolos (LVd + LVAd), que apresentam R^2 com significância a 0.001, porém, positivo.

Quadro 10 – Caracterização estatística dos índices derivados da imagem para as unidades de solo nas amostras de validação da classificação

Unidade	Clay mineral			Ferro mineral			Iron oxide			NDVI		
	Min	Med	Max	Min	Med	Max	Min	Med	Max	Min	Med	Max
PVAd	1,65	2,00	2,33	0,46	0,78	1,08	0,50	0,73	1,05	0,32	0,47	0,62
PVd + PVAd	1,70	2,20	2,68	0,45	0,57	0,92	0,45	0,56	0,73	0,37	0,57	0,67
CXbd1	1,38	1,58	1,82	0,91	1,24	1,57	0,65	0,91	1,37	0,00	0,17	0,28
CXbd	1,76	2,09	2,43	0,41	0,55	0,75	0,49	0,62	0,73	0,45	0,58	0,69
LVd + LVA	1,50	2,03	2,52	0,44	0,66	0,95	0,48	0,63	0,90	0,32	0,50	0,64
LVd + PVd	1,43	1,99	2,83	0,47	0,71	1,14	0,49	0,67	1,08	0,20	0,46	0,60
RLd + RRd	1,24	1,58	2,09	0,71	1,18	1,89	0,60	0,75	0,93	0,11	0,26	0,44
RYd	1,32	1,90	2,40	0,53	0,85	1,28	0,50	0,68	1,05	0,10	0,39	0,57
AR+ RLd	1,21	1,39	1,71	0,95	1,34	1,70	0,61	0,81	0,97	0,05	0,14	0,26

A correlação entre a variável CTI e a declividade para a maioria das unidades foi positiva e maior que 0,60, significativo a 1%. Valor negativo é observado, apenas, para a unidade de associação de Argissolos Vermelhos e Vermelho-Amarelos.

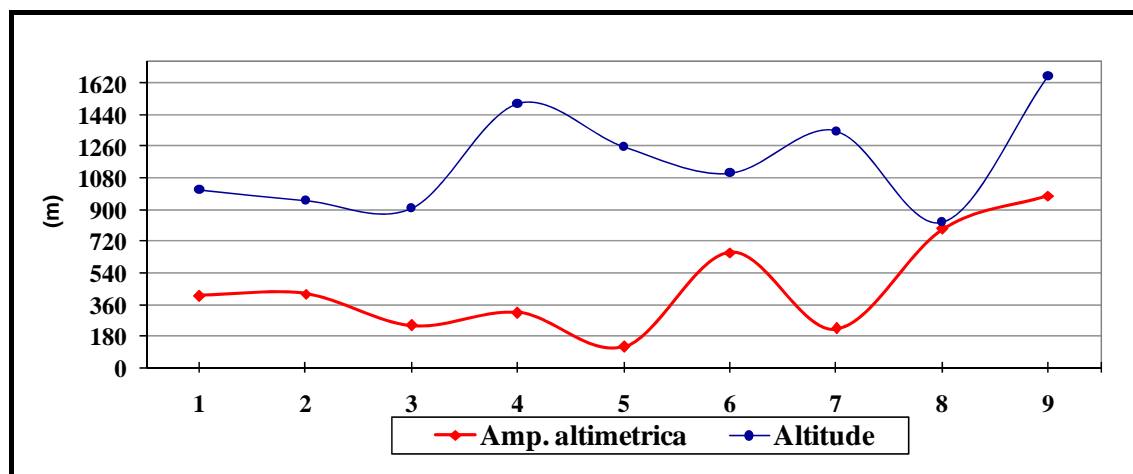


Figura 18 – Valores máximos das variáveis altitude e amplitude altimétricas nas amostras de validação das unidades de solo: 1))PVAd, 2)PVd + PVAd, 3)CXbd1, 4)CXbd, 5)LVd + PVd, 6)LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.

Quadro 11 – Correlação entre variáveis discriminantes para as amostras de treinamento das unidades de solos

Classe	Ampl.	Ferro mineral	Decliv.	Clay	Decliv.	Altitude	Decliv.
	Altitude	Iron oxide	Ampl.	NDVI	Radiação	Decliv.	CTI
PVAd	0,06 ^{NS}	0,06 ^{NS}	-0,11 ^{NS}	0,75	-0,25	0,26	0,45
PVd + PVAd	0,39	0,36	0,43	0,77	-0,15*	0,42	-0,82
CXbd1	0,32	0,38	-0,25	0,69	-0,52	-0,26	0,77
CXbd	0,54	-0,14*	0,46	0,65	0,04 ^{NS}	0,47	0,82
LVd+ PVd	0,05 ^{NS}	-0,27**	-0,11 ^{NS}	0,85	-0,39	0,12**	0,25
LVd+ LVAd	0,85	0,41**	0,40	0,83	0,39	0,38	0,70
RLd+RRd	0,66	-0,52	-0,77	0,70	-0,83	-0,78	0,77
RYd	-0,82	0,11 ^{NS}	0,72	0,79	-0,60	0,73	0,88
Ar + RLd	0,13*	-0,18	-0,01 ^{NS}	0,62	-0,66	0,13**	0,74
^(a) Total	0,64	-0,10 ^{NS}	0,17	0,92	-0,31	0,18	-0,67

Amp. (Amplitude altimétrica), Decliv. (declividade), Valores apresentados em negrito são significativos a 1%; **valores com 5% de significância; *valores com 10% de significância; ^{NS} valores não significativo a 10%; ^(a) correlação entre as 1.800 amostras.

4.1. Treinamento do classificador

A melhor arquitetura das redes para uma dada classificação depende da natureza dos dados discriminantes que variam com a escala e o tipo de levantamento, do nível de ruído e também da complexidade das classes mapeadas. Assim, embora existam na literatura algumas indicações de configurações de arquitetura de rede definidas de forma experimental, essa é ainda melhor definida para cada classificação específica por tentativas e erros (KAVZOGLU & MATHER, 2003).

Os testes para escolha da arquitetura da rede permitiram definir o número de neurônios na camada interna das redes, de ciclo de treinamento, e de variáveis discriminantes para discriminar os solos na classificação pelas redes neurais.

Na classificação pelo MaxVer, a contribuição das variáveis discriminantes foi determinada da mesma maneira como feita para a classificação pelas redes neurais.

4.1.1. Número de neurônios

A classificação foi iniciada com um neurônio na camada interna, sendo esse número aumentado de um a um, de acordo com o método construtivo (HIROSE et al., citados por KAVZOGLU & MATHER, 1999). Os neurônios foram adicionados até o total de 15, quando não houve mais aumento na exatidão da classificação, sendo esse o número máximo utilizado. O resultado das classificações, o índice kappa e a variância, das 15 classificações realizadas com diferentes arquiteturas internas foram avaliadas pelo teste de significância estatística (Quadro 12).

Na matriz de significância, os valores da diagonal principal representam a exatidão do classificador em relação a uma classificação aleatória, enquanto que os valores de fora da diagonal da matriz representam o desempenho relativo entre os classificadores (VIEIRA, 2000). Nessa matriz, o resultado do teste Z apresentado na diagonal principal mostram que todas as classificações obtiveram um resultado melhor do que o de uma classificação aleatória. O índice kappa aumentou de forma linear com o aumento do número de neurônios até o total de nove.

O menor índice kappa, com exatidão de 0,33, foi obtido para a classificação com um neurônio, enquanto que o maior índice ($kappa = 0,948$) ocorreu na classificação com nove neurônios. A arquitetura com nove neurônios foi considerada como a mais indicada para a classificação dos solos com as variáveis utilizadas, embora o resultado com essa arquitetura tenha sido idêntico ao apresentado com 12 neurônios e não tenha diferido estatisticamente das classificações com 10, 11, 12 e 14 neurônios. Nesse caso, foi escolhida a rede de menor complexidade, já que o aumento na estrutura das redes resultou em diferença pouco significativa na exatidão da classificação, e já que esse é um parâmetro de menor impacto nos resultados (FOODY & ARORA, 1997).

Foody e Arora (1997) e Vieira (2000) colocam que a escolha da melhor arquitetura e parâmetro da classificação deve ser feita a partir de treinamentos e avaliação dos resultados para cada projeto executado, já que diferenças no ambiente e no tipo de variáveis utilizadas para áreas distintas levam também a diferentes resultados.

O melhor desempenho do classificador, em relação ao número de neurônios, no presente estudo, foi semelhante ao encontrado por Chagas (2005), o qual obteve o melhor resultado utilizando entre cinco e oito neurônios na camada interna das redes durante a classificação de solos em duas áreas de diferentes formações litológicas.

Quadro 12 – Matriz de significância para número de neurônios na camada interna da rede

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kappa	0,329	0,659	0,731	0,897	0,923	0,918	0,925	0,936	0,948	0,946	0,941	0,948	0,93	0,943	0,918
¹ Var. *(10) ⁻³	0,014	0,147	0,126	0,058	0,045	0,047	0,044	0,038	0,031	0,032	0,035	0,031	0,029	0,034	0,047
1	88,564														
2	26,024*	54,353													
3	33,999*	4,358*	65,123												
4	67,033*	16,623*	12,238*	117,782											
5	77,464*	19,053*	14,683*	2,562*	137,593										
6	75,538*	18,595*	14,217*	2,049*	0,521	133,904									
7	78,394*	19,247*	14,879*	2,772*	0,212	0,734	139,449								
8	84,338*	20,365*	16,008*	3,98*	1,427	1,952	1,215	151,839							
9	92,481*	21,661*	17,318*	5,406*	2,868*	3,397*	2,656*	1,445	170,266						
10	91,17*	21,451*	17,104*	5,165*	2,621*	3,15*	2,409*	1,195	0,252	167,231					
11	87,608*	20,903*	16,55*	4,563*	2,012*	2,54*	1,8	0,585	0,862	0,611	159,058				
12	92,481*	21,661*	17,318*	5,406*	2,868*	3,397*	2,656*	1,445	0	0,252	0,862	170,266			
13	91,866*	20,427*	15,984*	3,538*	0,814	1,376	0,585	0,733	2,324*	2,049*	1,375	2,324*	172,697		
14	88,808*	21,11*	16,76*	4,796*	2,25*	2,778*	2,038*	0,825	0,62	0,369	0,241	0,62	1,638	161,723	
15	75,538*	18,595*	14,217*	2,049*	0,521	0	0,734	1,952	3,397*	3,15*	2,54*	3,397*	1,376	2,778*	133,904

¹ Variância; * Existe diferença estatística ao nível de 95% de probabilidade. Z Tab. = 1,96. Todas as classificações foram realizadas com 10 mil ciclos de treinamento.

4.1.2. Ciclos de treinamento na classificação pelas redes neurais

A avaliação do classificador no que se refere ao desempenho apresentado em função do número de ciclos de treinamento foi realizada utilizando-se a mesma arquitetura das redes para todas as classificações. Na camada de entrada utilizaram-se todas as 19 variáveis discriminantes, na camada de saída utilizaram-se as nove unidades de solo definidas pelas associações de classes de solos e, conforme definido pela exatidão do índice kappa, foram utilizados nove neurônios na camada interna.

O número inicial de ciclos de treinamento foi determinado de forma heurística, iniciando-se a classificação com um número de cinco mil. Esse número foi aumentando de cinco em cinco mil até o total de 30 mil. Realizaram-se seis classificações, cujos resultados foram submetidos a um teste de significância estatística para se observar a diferença em função do número de ciclos de treinamento (Quadro 13).

Na matriz de significância observa-se, na diagonal principal, que os maiores valores do teste ocorreram para as classificações com 10 e 30 mil ciclos de treinamento, não havendo diferença estatística significativa entre as duas. Assim, embora a classificação com 30 mil ciclos tenha sido ligeiramente superior à obtida com 10 mil, selecionou-se a menor arquitetura, ou seja, com 10 mil ciclos, para a realização das classificações posteriores empregadas na escolha das variáveis discriminantes.

Quadro 13 – Matriz de significância para o número de ciclos de treinamento

Ciclos	5.000	10.000	15.000	20.000	25.000	30.000
Kappa	0,937	0,948	0,932	0,947	0,939	0,951
Variância	0,000037	0,000031	0,00004	0,000032	0,000036	0,00003
5.000	154,042					
10.000	1,334	170,266				
15.000	0,57	1,899	147,362			
20.000	1,204	0,126	1,768	167,408		
25.000	0,234	1,1	0,803	0,97	156,5	
30.000	1,71*	0,384	2,271*	0,508	1,477	173,628

* Existe diferença estatística ao nível de 95% de probabilidade; Z Tab. = 1,96. Classificação feita com arquitetura interna da rede composta por 19 variáveis discriminantes e nove neurônios.

A determinação da arquitetura baseou-se na exatidão apresentada, considerando-se também a estrutura das redes e o tempo gasto no treinamento. A rede com menor número de neurônios foi selecionada, visando obter maior número de acertos na

classificação, além do fato de a mesma demandar um menor tempo no treinamento do classificador.

4.1.3. Avaliação das variáveis discriminantes

A contribuição das variáveis discriminantes na classificação pelas redes neurais e pelo MaxVer foi avaliada de forma interativa, removendo-se uma variável por vez, ou um conjunto de variáveis, do conjunto de todas as 19 selecionadas para a análise discriminatória dos solos. As seis cenas da imagem de satélite foram consideradas em conjunto como uma única variável de análise. A partir do conjunto composto por todas as variáveis formaram-se 13 subconjuntos, sendo 12 deles com 18 variáveis cada, e um, sem as imagens, com 13 variáveis. Os valores do índice kappa e da variância foram utilizados como indicadores de exatidão, para se comparar o desempenho do classificador em função da contribuição das variáveis de cada um dos 13 subconjuntos classificados por cada um dos dois classificadores. Os resultados obtidos na classificação foram submetidos à análise pelo Teste Z, a fim de se observar a existência de diferença significativa entre os conjuntos de variáveis utilizadas.

4.1.4. Variáveis discriminantes na classificação pelas redes neurais

A classificação feita com todas as 19 variáveis foi utilizada como referência para as comparações entre conjuntos de variáveis. Essa classificação apresentou índice kappa de 0,948 (Quadro 15). Nessa matriz, observa-se que todas as unidades de solo apresentaram índice de exatidão maior que 80%. Outros 14 conjuntos de variáveis foram classificados, e os resultados comparados com o conjunto de referência pelo Teste Z, com a finalidade de se avaliar a significância estatística na exatidão das classificações devido a utilização de diferentes conjunto de variáveis (Quadro 17).

As variáveis geologia, radiação solar, declividade e amplitude altimétrica, e os índices, *iron oxide* e CTI contribuíram de forma significativa para a discriminação dos solos. A exatidão diminuiu em até 8% quando essas variáveis foram excluídas do conjunto, uma por vez antes da classificação. O índice CTI contribuiu em menor proporção, com 1%; o índice *iron oxide* contribuiu com 3%, e com 3,7% a amplitude

altimétrica; já a geologia, com 6%, e com 8%, a radiação solar. Por outro lado, o mapa de solos, o MDE, as bandas da imagem de satélite e o índice *Fe minerals* contribuíram para aumentar a exatidão, mas não apresentaram diferenças estatísticas.

A exatidão do mapa não foi alterada quando se suprimiu do conjunto de variáveis o índice NDVI. Esse índice apresentou pequena variação de valores entre as unidades de solo, conforme apresentado no quadro 10 e na figura 19. A exatidão do mapa não sofreu alteração com a eliminação do índice, no entanto, houve diferenças na exatidão de algumas unidades de solo. Os Argissolos e Latossolos, que geralmente ocorrerem em áreas onde a cobertura vegetal é mais densa e difere daquelas dos Campos Rupestres e Campos Cerrados das áreas vizinhas, tiveram maior exatidão quando o NDVI foi utilizado. Conforme se observa nas matrizes de confusão das classificações feitas com e sem o índice, enquanto a exatidão dos Argissolos e Latossolos diminuiu, ao eliminar o NDVI, a dos Neossolos Litólicos aumentou em 13% (Quadros 15 e Quadro 1 - Anexo).

Moran e Bui (2002) e Dobos et al. (2000) observaram maior exatidão no mapeamento de solos com o uso do índice NDVI. Os primeiros utilizaram dados espectrais do sensor MSS do Landsat, para geração do índice; já os segundos, utilizando imagem AVHRR do satélite NOAA, conseguiram discriminar os solos com 87% de exatidão, quando utilizaram em conjunto com o índice a geologia e os atributos do terreno. Por outro lado, assim como ocorreu nesse estudo, Scull et al. (2005), em trabalho de mapeamento de solos de uma área desértica, empregaram o modelo de árvore de decisão e não observaram contribuição significativa do índice NDVI.

A geologia apresentou a terceira maior contribuição para discriminar os solos. A importância dessa variável relaciona-se com o fato do classificador de redes neurais não assumir uma distribuição inicial para os dados. Aliado a isso, ocorre que as características dos solos na área apresentam forte relação pedogenética com a itologia da área, a exemplo dos Neossolos Litólicos, que ocorrem sob os quartzitos, dos Neossolos Flúvicos, sob sedimentos, e dos Latossolos e Argissolos de calcários e gnaisses. A contribuição da geologia na discriminação de solos foi também observada por Scull et al (2005), ao verificarem que das 32 variáveis utilizadas, a geologia juntamente com atributos do terreno foram as mais importantes.

Quadro 14 – Matriz de significância para as classificações feitas com 15 diferentes conjuntos de variáveis pelas redes neurais

Conjunto	Todas 19 variáveis	Geologia	Solos	Imagem	Clay mineral	Ferro mineral	Iron oxide	NDVI	Aspecto	Curvatura	MDE	Declividade	CTI	Radiação	Ampl. altimétrica
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kappa	0,948	0,889	0,939	0,94	0,964	0,937	0,918	0,948	0,951	0,952	0,936	0,881	0,939	0,868	0,911
¹ Var. *(10) ⁻³	0,031	0,062	0,036	0,035	0,022	0,037	0,047	0,031	0,029	0,029	0,038	0,066	0,036	0,072	0,051
1	170,266														
2	6,118*	112,903													
3	1,1	5,051*	156,5												
4	0,985	5,178*	0,119	158,889											
5	2,198*	8,183*	3,283*	3,179*	205,525										
6	1,334	4,824*	0,234	0,354	3,515*	154,042									
7	3,397*	2,778*	2,305*	2,429*	5,538*	2,073*	133,904								
8	0	6,118*	1,1	0,985	2,198*	1,334	3,397*	170,266							
9	0,387	6,499*	1,488	1,375	1,82	1,723	3,785*	0,387	176,596						
10	0,516	6,604*	1,612	1,5	1,68	1,846	3,9*	0,516	0,131	176,782					
11	1,445	4,7*	0,349	0,468	3,615*	0,115	1,952	1,445	1,833	1,955	151,839				
12	6,803*	0,707	5,743*	5,871*	8,848*	5,518*	3,481*	6,803*	7,182*	7,284*	5,393*	108,444			
13	1,1	5,051*	0	0,119	3,283*	0,234	2,305*	1,1	1,488	1,612	0,349	5,743*	156,5		
14	7,883*	1,814	6,832*	6,961*	9,902*	6,609*	4,583*	7,883*	8,259*	8,358*	6,484*	1,107	6,832*	102,295	
15	4,086*	2,07*	3,002*	3,127*	6,203*	2,772*	0,707	4,086*	4,4720*	4,584*	2,65*	2,774*	3,002*	3,877*	127,566

¹ Variância; Z Tab. = 1,96. * Existe diferença estatística ao nível de 95% de probabilidade. Conjuntos de variáveis: 1)todas as 19 variáveis. Conjuntos de 2 a 15 (eliminação de uma das variáveis do conjunto de 19): 2) sem geologia, 3) sem mapa de solo, 4) sem bandas das imagens de satélite, 5) sem o índice *Clay minerals*, 6) sem o índice *Ferrous minerals*, 7) sem o índice *Iron oxide*, 8) sem o índice NDVI, 9) sem aspecto, 10) sem curvatura 11) sem MDE, 12) sem declividade 13) sem CTI, 14) sem radiação solar, 15) sem amplitude altimétrica.

A variável aspecto não apresentou contribuição significativa para separar as unidades de solo definidas neste estudo. A pouca contribuição apresentada por essa variável pode decorrer da relação entre essa com outras, como a curvatura, a declividade e mesmo o mapa de altitude, já que a distribuição dos solos nas classes da variável aspecto mostra certa heterogeneidade (Figura 20).

Observa-se que os Cambissolos de Floresta Nebular apresentaram maior homogeneidade; as demais unidades tiveram predominância de uma ou outra classe. As unidades CXbd e RLd apresentaram maior proporção de área com orientação para leste, enquanto que as unidades de Argissolos estão mais orientadas para sul. A unidade CXbd1, por sua vez, apresenta certa homogeneidade entre as quatro classes. Embora a contribuição de tal variável não tenha apresentado diferença significativa na exatidão global, houve diferença na discriminação dos Latossolos e Argissolos.

Enquanto os Latossolos e Argissolos apresentaram maior exatidão quando a variável aspecto foi incluída na classificação, os Neossolos (RLd + RRd) tiveram menor índice de acerto, conforme pode ser observado nas matrizes de confusão das classificações feitas com e sem a variável aspecto (Quadros 16 e Quadro 2 - Anexo).

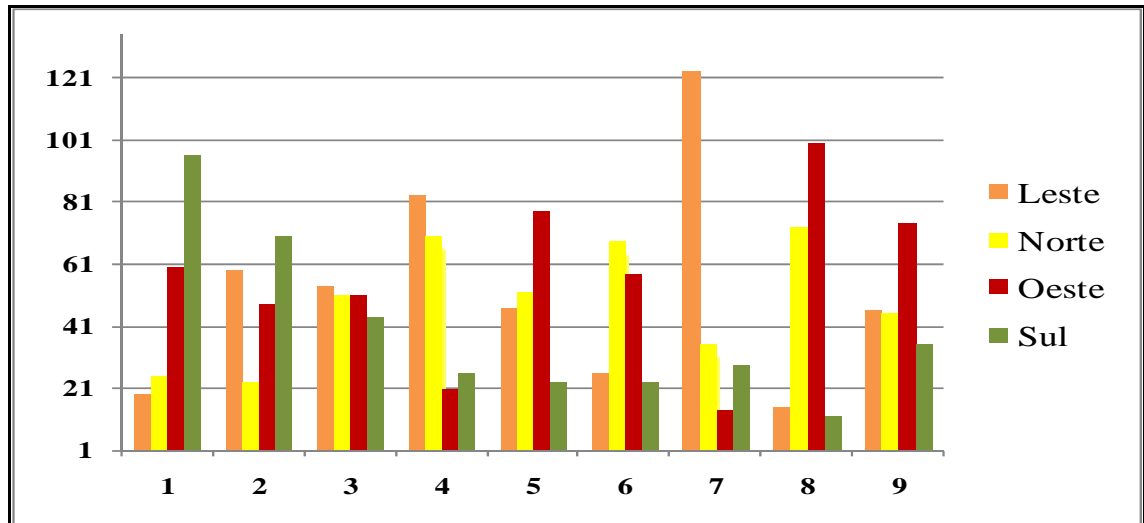


Figura 20 – Frequência da variável aspecto nas amostras de validação para as unidades de solo: 1)) PVAd, 2)PVd + PVAd, 3) CXbd1, 4) CXbd, 5) LVd + PVd, 6) LVd + LVA, 7) RLd + RRd, 8) RYd, 9)AR + RLd.

O melhor desempenho na classificação pelo MaxVer ocorreu quando foi excluído do conjunto com 19 variáveis o índice *Clay minerals*, obtendo exatidão de 0,96 pelo índice kappa (Quadro 16). Esse valor foi ligeiramente superior ao obtido na classificação com todas as variáveis, para a qual se obteve kappa de 0,948 (Quadro 2).

Essas duas classificações apresentaram índices de concordância considerados bom, conforme Landis e Koch (citados por Congalton e Green, 1999).

Na matriz de confusão da classificação sem o índice *Clay mineral*, pode-se observar uma diferença de 22%, entre a maior e a menor exatidão obtida para as unidades de solo, sendo a unidade de Neossolo, aquela classificada com o menor índice (84%), enquanto que o maior índice foi obtido pela unidade de LVA + PVAd, com 90,7% das amostras corretamente classificadas (Quadro 17).

Quadro 15 – Matriz de confusão para a classificação feita pelas redes neurais com todas as 19 variáveis

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão Usuário	Z
PVAd	192	0	0	1	7	0	0	0	0	200	96	61.554
PVd + PVAd	0	195	1	0	0	0	0	0	0	196	99,5	173.768
CXbd1	0	0	198	1	0	1	1	0	0	201	98,5	102.368
CXbd	0	0	0	197	4	0	10	1	1	213	92,5	45.488
LVd + PVd	6	0	0	0	184	1	11	4	0	206	89,3	36.822
LVd + LVA	0	5	0	0	0	198	3	0	0	206	96,1	63.449
RLd + RRd	2	0	1	1	4	0	160	0	0	168	95,2	51.446
RYd	0	0	0	0	1	0	0	195	1	197	99	123.18
AR + RLd	0	0	0	0	0	0	15	0	198	213	93	47.104
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	96	97,5	99	98,5	92	99	80	97,5	99			

Kappa = 0,948; Exatidão global = 95,4; Variância = 0,000031; Z estatístico = 170,560, Z Tab = 1,96.

Quadro 16 – Matriz de confusão para a classificação feita pelas redes neurais sem a variável *Clay mineral*

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exat. Usuário	Z
PVAd	196	2	1	0	6	7	2	0	0	214	91,6	42,872
PVd + PVAd	0	195	0	0	4	7	0	0	0	206	94,7	53,686
CXbd1	0	0	199	2	0	0	0	0	0	201	99	125,697
CXbd	0	0	0	197	3	0	1	0	2	203	97	72,55
LVd + PVd	4	0	0	0	177	0	0	1	0	182	97,3	71,307
LVd + LVA	0	3	0	0	3	186	0	0	0	192	96,9	68,543
RLd + RRd	0	0	0	0	2	0	197	0	1	200	98,5	101,854
RYd	0	0	0	1	5	0	0	199	1	206	96,6	68,005
AR + RLd	0	0	0	0	0	0	0	0	196	196	100	0
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	98	97,5	99,5	98,5	88,5	93	98,5	99,5	98			

Kappa = 0,964; Exatidão global = 96,8; Variância = 0,000022; Z estatístico = 205,850, Z Tab = 1,96.

4.1.5. Variáveis discriminantes na classificação pelo MaxVer

As variáveis discriminantes utilizadas na classificação pelo algoritmo MaxVer foram selecionadas utilizando-se o mesmo critério empregado na classificação pelas redes neurais, ou seja, eliminando-se do conjunto total de variáveis uma variável por vez, ou um conjunto de variáveis semelhantes. Na avaliação dos resultados pelo Teste Z, compararam-se a exatidão obtida na classificação do conjunto de referência, composto por todas as variáveis discriminantes, e a exatidão dos demais subconjuntos formados.

A primeira classificação utilizada como referência para avaliar o desempenho das variáveis foi realizada utilizando-se o conjunto com todas as 19 variáveis. Essa classificação apresentou exatidão global de 41,6%, com três unidades de solos discriminadas com índice acima de 90%, uma unidade com 78% e cinco unidades com todas as amostras incorretamente classificadas (Quadro 3 - Anexo). O resultado dessa classificação foi comparado com aqueles obtidos por 13 outras classificações feitas com os subconjuntos dos quais foi eliminada uma das variáveis de análise.

Verificou-se que a geologia respondeu por 43% dos erros do classificador. Ao suprimir essa variável, o índice kappa passou de 0,34 para 0,79. Na classificação sem a geologia, com exceção dos Neossolos Flúvicos, todas as demais unidades foram discriminadas com mais de 90% de exatidão (Quadro 4 - Anexo). A partir dessa constatação, retirou-se a geologia dos subconjuntos de variáveis, os quais foram novamente submetidos à classificação. Os resultados foram novamente avaliados pelo teste de significância estatística, mostrando que a curvatura apresentou a menor contribuição, elevando o erro em 7,5% em relação ao conjunto de referência.

Na sequência, os subconjuntos formados sem as variáveis geologia e curvatura tiveram uma das demais variáveis excluídas e foram submetidos a nova classificação. Dois conjuntos adicionais foram formados, sendo um conjunto sem os três índices (*Clay mineral*, *Iron oxide* e *Ferro mineral*) e outro conjunto sem as informações da imagem espectral, ou seja, sem as bandas da imagem de satélite e sem os quatro índices dela derivados. No quadro 17 são apresentados os resultados do teste de significância entre a classificação de referência e os 15 conjuntos de variáveis classificados.

As variáveis que apresentaram diferença estatística significativa na classificação dos solos pelo algoritmo MaxVer foram: *Clay mineral*, NDVI, CTI e declividade. A

imagem não contribuiu. A melhor classificação foi obtida quando se eliminou do conjunto total as seguintes variáveis: geologia, curvatura e imagens. A exatidão obtida apresentou alto grau de concordância, com índice kappa de 0,89. Esse conjunto de variáveis foi submetido a todos os processos de classificação para a geração do mapa.

O conjunto de variáveis discriminantes que apresentou a maior exatidão na classificação dos solos pelo MaxVer foi composto por 11 variáveis, quais sejam: MDE, aspecto, CTI, declividade, diferença altimétrica, radiação solar, mapa de solos e índices derivados da imagem (ferro mineral, clay mineral, iron oxide e NDVI), sendo excluídas da classificação as variáveis: geologia, curvatura e das bandas da imagem de satélite.

Quadro 17 – Matriz de significância entre as classificações feitas com 15 conjuntos de variáveis discriminantes pelo MaxVer

Conjunto	Geologia/ curvatura	Imagem	Solos	Clay	Fe	Iron	NDVI	Aspecto	CTI	Declividade	Dif. Altimétrica	MDE	Radiação	Índices	espectral
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kappa	0,866	0,889	0,844	0,68	0,871	0,878	0,791	0,872	0,783	0,762	0,849	0,744	0,858	0,887	0,754
¹ Var. *(10) ⁻³	0,073	0,062	0,083	0,073	0,071	0,068	0,103	0,106	0,106	0,114	0,081	0,12	0,078	0,063	0,119
1	101,358														
2	1,98*	112,903													
3	1,761	3,737*	92,641												
4	15,393*	17,988*	13,131*	79,588											
5	0,417	1,561	2,176*	15,917*	103,369										
6	1,011	0,965	2,767*	16,675*	0,594	106,473									
7	5,653*	7,629*	3,886*	8,367*	6,065*	6,653*	77,94								
8	0,448	1,312	2,037*	14,351*	0,075	0,455	5,603*	84,696							
9	6,204*	8,178*	4,437*	7,699*	6,614*	7,202*	0,553	6,113*	76,052						
10	7,605*	9,573*	5,842*	5,996*	8,014*	8,598*	1,969*	7,416*	1,416	71,368					
11	1,37	3,345*	0,39	13,618*	1,784	2,376*	4,276*	1,682	4,826*	6,23*	94,333				
12	8,782*	10,748*	7,019*	4,607*	9,189*	9,773*	3,147*	8,514*	2,594*	1,177	7,406*	67,918			
13	0,651	2,62*	1,103	14,485*	1,065	1,655	4,98*	1,032	5,529*	6,928*	0,714	8,102*	97,149		
14	1,801	0,179	3,559*	17,75*	1,382	0,786	7,451*	1,154	8*	9,396*	3,167*	10,571*	2,442*	111,751	
15	8,083*	10,034*	6,332*	5,34*	8,488*	9,068*	2,483*	7,867*	1,933	0,524	6,718*	0,647	7,41*	9,859*	69,119

¹ Variância; * Existe diferença estatística ao nível de 95% de probabilidade; Z Tab. = 1,96; O conjunto 1 foi composto por 17 variáveis, sendo excluídos do conjunto de 19 as variáveis geologia e curvatura; a partir desse conjunto compuseram-se os conjuntos de 2 a 15, sendo excluídas uma ou mais variáveis de cada um, como se segue: 2- sem bandas das imagens ETM+, 3- sem mapa de solo, 4- sem *Clay minerals*, 5- sem *Ferrous minerals*, 6- sem *Iron oxide*, 7- sem NDVI, 8- sem aspecto, 9- sem CTI, 10- sem declividade, 11- sem amplitude altimétrica, 12- sem MDE, 13- sem radiação solar, 14- sem índices (*Clay minerals*, *Ferrous minerals* e *Iron oxide*). 15- sem dados espectrais da imagem (bandas da imagem e índices derivados).

Quadro 18 – Matriz de confusão gerada na classificação pelo MaxVer com 11 variáveis

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão Usuário
PVAd	185	12	0	0	12	0	0	10	0	219	84,5
PVd + PVAd	4	178	0	0	4	34	0	1	0	221	80,5
CXbd1	0	2	195	0	0	0	0	0	0	197	99
CXbd	0	0	0	198	0	0	1	0	0	199	99,5
LVd + PVd	10	2	4	0	156	0	0	0	1	173	90,2
LVd + LVA	1	6	0	0	0	166	0	40	0	213	77,9
RLd + RRd	0	0	1	2	28	0	199	0	2	232	85,8
RYd	0	0	0	0	0	0	0	149	0	149	100
AR + RLd	0	0	0	0	0	0	0	0	197	197	100
Total	200	200	200	200	200	200	200	200	200	1800	
Exatidão Prod.	92,5	89	97,5	99	78	83	99,5	74,5			

Kappa = 0,889; Exatidão global = 90,2; Variância = 0,000062; Z calculado = 112,752; Z tabelado = 1,96, Variáveis utilizadas: (conjunto de 19 variáveis, sem a geologia, a curvatura e as seis imagem de satélite).

4.2. Comparação entre as classificações feitas pelas Redes Neurais e pelo MaxVer

As duas classificações que apresentaram os maiores índices de exatidão pelas redes neurais foram comparadas entre si e com a classificação obtida com maior índice de acerto pelo MaxVer. Os índices kappa, com as respectivas variâncias de cada uma dessas classificações, foram usados para verificar diferença significativa por meio do Teste Z. O resultado do teste mostra que a 5% de significância há diferença entre todas as classificações avaliadas (Quadro 19).

Quadro 19 – Matriz de significância para as melhores classificações obtidas pelas redes neurais e pelo MaxVer no treinamento do classificador

Classificador	Redes Neurais (19variáveis)	Redes Neurais (18variáveis)	Máxima Verossimilhança
Kappa	0,948	0,964	0,878
Variância	0,000031	0,000022	0,000068
Redes Neurais (19)	170,266		
Redes Neurais (18)	2,198*	205,525	
MaxVer	7,035*	9,065*	106,473

* Há diferença estatística ao nível de 95% de probabilidade.

As variáveis geologia e curvatura não contribuíram com a classificação dos solos pelo MaxVer. Essas variáveis são categóricas e não apresentam valores com distribuição que possam definir uma função com distribuição normal, na qual o classificador baseia-se para fazer a classificação. Por outro lado, na classificação pelas redes neurais, a variável geologia, quando suprimida, apresentou para todos os conjuntos testados maior erro de classificação. A variável aspecto não apresentou diferença estatística quando utilizada na classificação pelos dois classificadores.

Os dados espectrais das imagens foram importantes na classificação pelos dois modelos de classificação avaliados. No MaxVer, esses foram empregados na forma de índices, enquanto que nas redes neurais, além dos índices, utilizaram-se as bandas e as imagens. Dos índices derivados da imagem, o NDVI teve contribuição significativa pelo MaxVer, enquanto que pelas redes neurais, sua contribuição não foi significativa. O índice *Clay mineral*, na avaliação do treinamento, levou a um aumento no confundimento do classificador, quando utilizado na classificação pelas redes neurais.

Os atributos do terreno, declividade, CTI e altitude, por serem variáveis com valores contínuos, tiveram contribuição significativa na discriminação dos solos em ambos os modelos: redes neurais e MaxVer.

A diferença nos resultados obtidos com as variáveis categóricas e contínuas pelo algoritmo mostram como o MaxVer e as redes neurais diferem em relação ao uso de informações com diferentes padrões de distribuição.

4.3. Classificação e pós-processamento dos mapas de solos

Ao final do treinamento do classificador de redes neurais realizou-se a predição dos solos para toda a área em estudo. O processo foi executado atribuindo-se a cada pixel da imagem o identificador da unidade de solo ao qual ele apresentou a maior probabilidade de pertencer, conforme os padrões informados no treinamento.

Os mapas gerados foram pós-processados para eliminação de ruídos representados por pequenas áreas sem expressão cartográfica na escala utilizada. O procedimento seguiu o recomendado em EMBRAPA (1997), para a área mínima representável, conforme a escala de confecção do mapa, adotando o limite de 10 ha. Assim, as áreas com dimensão menor do que 10 ha foram reclassificadas, recebendo a informação da unidade de solo vizinha. Os processamentos foram realizados no programa ArcGis 9.2, utilizando-se as ferramentas do módulo *spatial analyst*.

Os mapas de solos, obtidos após a classificação e pós processamentos, são apresentados nas figuras 21 e 22, respectivamente para a classificação feita pelas redes neurais, utilizando todas as variáveis, e pela classificação MaxVer, com 11 variáveis.

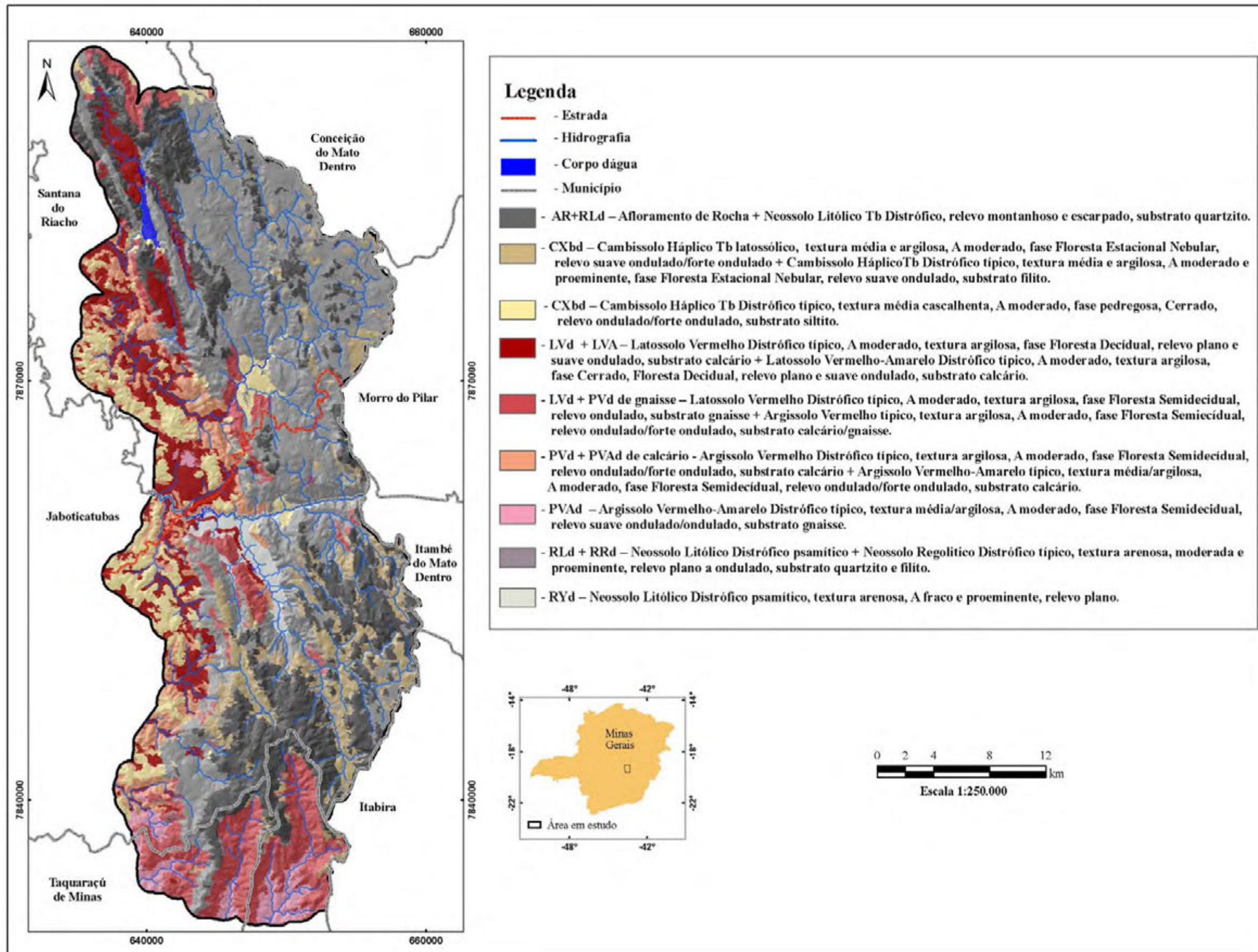


Figura 21 – Mapa de solos obtido por classificação pelas Redes Neurais Artificiais.

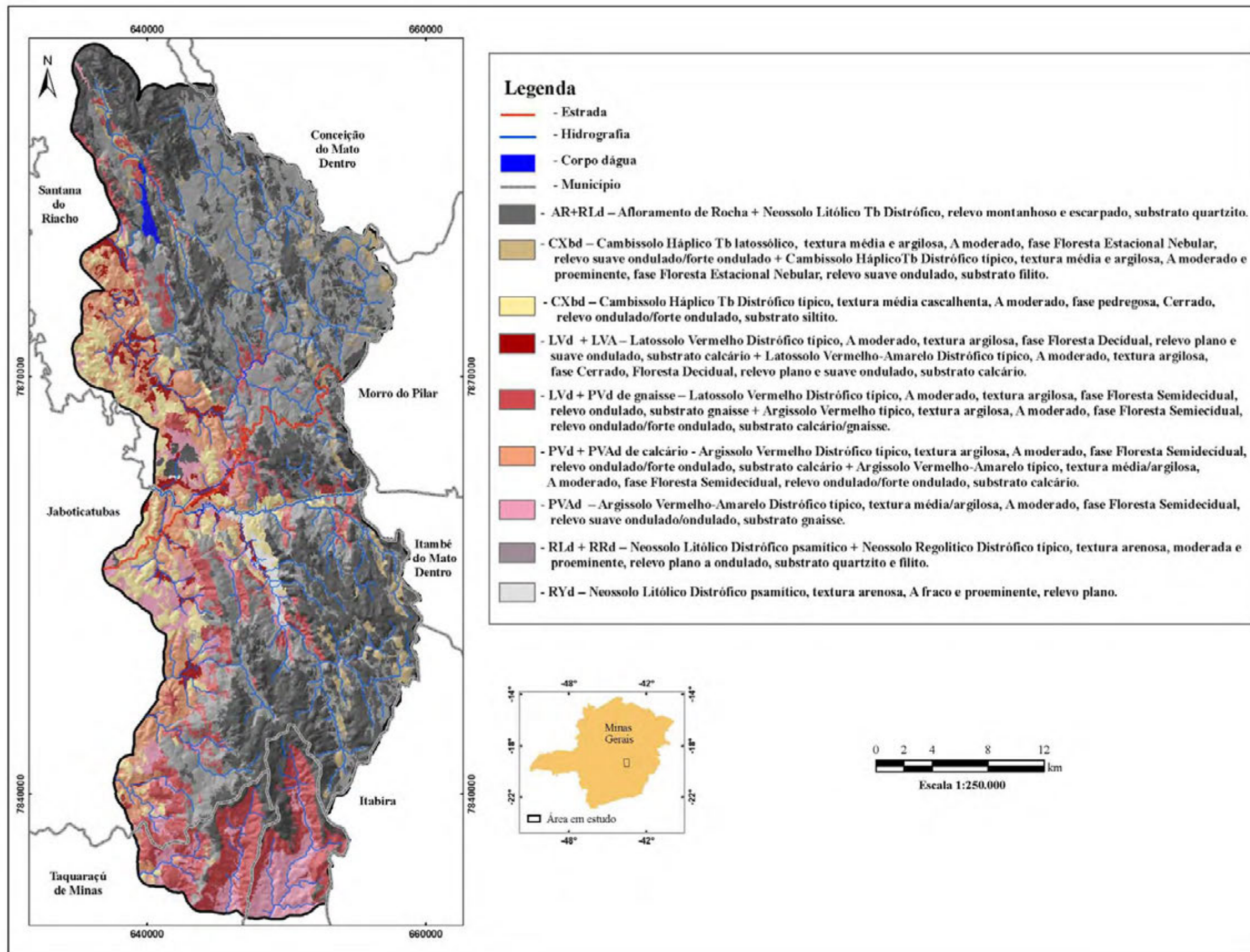


Figura 22 – Mapa de solos obtido por classificação pelo MaxVer.

4.4. Validação do mapa de solos com pontos de referência terrestre

A validação dos mapas gerados pelos classificadores de redes neurais, pelo MaxVer e pelo método convencional foi realizada utilizando-se os 362 pontos de referência terrestre coletados de forma aleatória. A concordância foi avaliada pelos índices kappa e de exatidão global, conforme proposto por Congalton e Green (1999). Nesse procedimento empregou-se a técnica de álgebra de mapas, utilizando-se o modelo raster para a validação dos mapas obtidos com os classificadores supervisionados; já para o mapa convencional, foi utilizado o modelo vetorial.

Os mapas validados para a abordagem pelas redes neurais foram aqueles que apresentaram os maiores acertos no treinamento: o mapa gerado com o conjunto sem a variável *Clay mineral* e o mapa gerado com o conjunto composto por todas as variáveis. Na abordagem pelo MaxVer, também foi validado o mapa melhor classificado.

As matrizes de confusão geradas na validação dos três mapas são apresentadas nos quadros 6, 7 e 8 (Anexo), respectivamente para as classificações feitas pelas redes neurais com 19 e 18 variáveis e para o classificação pelo MaxVer. A comparação dos resultados dessas classificações pelo Teste Z é apresentada no quadro 20.

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 inferiores aos obtidos pelas redes com 18 variáveis e superiores aos das redes com 19 variáveis. A exatidão global do MaxVer foi de 59% (kappa = 0,547).

Quadro 20 – Matriz de significância para a validação com pontos de referência terrestre para as classificações obtidas pelas Redes Neurais e pelo MaxVer

Classificador	Redes Neurais (19 variáveis)	Redes Neurais (18 variáveis)	MaxVer (11 variáveis)
Kappa	0,508	0,564	0,547
Exatidão global	54,3	60,1	59,0
Variância	0,00092	0,000898	0,000869
Redes Neurais 18	16,748		
Redes Neurais 19	1,313	18,821	
MaxVer	0,922	0,404	18,556

* Não há diferença significativa a 95% de probabilidade. Classificação pelas redes neurais feita utilizando todas as 19 variáveis e utilizando 18 variáveis (sem a variável *Clay mineral*); classificação pelo MaxVer utilizando 11 variáveis (sem as variáveis: geologia, curvatura e bandas da imagem).

Os resultados do teste Z mostram que a exatidão dos três mapas não apresentou diferença estatística significativa com 5% de significância. Assim, embora na validação com amostras de treinamento para as classificações feitas pelas redes neurais, o resultado do teste Z tenha mostrado diferença significativa, o mesmo não ocorreu quando os mapas foram validados com pontos de referência.

Os resultados da validação com amostras de treinamento e com pontos de referência terrestre mostram que esses são independentes, para ambos os classificadores e conjuntos de variáveis. No treinamento, no caso das redes neurais, é feito o reconhecimento dos padrões e, posteriormente, o mesmo é generalizado para todos os valores do mapa. Assim, as escolhas tanto das variáveis discriminantes, quanto do classificador devem ser feitas tendo como base os resultados da validação dos dois processos, treinamento e validação.

Chagas (2006), obteve exatidão global de 70,8% na classificação de solos pelas redes neurais e, 50,7% pelo MaxVer, utilizando dados em escala de 1:50.000. No presente estudo a concordância foi superior para o MaxVer e inferior em 18% para as abordagens por redes neurais. As diferenças entre os resultados obtidos e aqueles do referido trabalho devem considerar fatores como escala de mapeamento, variáveis utilizadas e complexidades dos solos mapeados, os quais dificultam uma comparação direta entre a exatidão obtida dos mapeamentos.

4.4.1. Avaliação das unidades de solos classificadas pelo MaxVer e pelas Redes Neurais

As unidades de solos classificadas pelas redes neurais e pelo MaxVer foram comparadas a partir dos resultados dos mapas que obtiveram os maiores índices de acerto na validação com pontos de referência terrestre. O índice kappa (kappa condicional) e a variância de cada uma das unidades de solo foram utilizados no teste Z para avaliar a existência de diferenças estatísticas entre os classificadores.

Os resultados do teste de significância estatística para a classificação pelo MaxVer e pelas redes neurais com 18 e 19 variáveis estão apresentados no quadro 21. Em conjunto, os resultados de significância e os valores de índice kappa e variância obtidos para cada uma das unidades de solo (Quadro 22) mostram que: o MaxVer classificou duas unidades com exatidão estatisticamente superior à obtida na

classificação pelas redes neurais (RRL + RRd e AR + RLd). Já as redes neurais classificaram quatro unidades de solo com exatidão estatisticamente superior à apresentada pelo MaxVer (CXbd, RY, PVA, LVd + LVA). Três unidades não tiveram diferenças estatísticas significativas 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 diferenças estatísticas significativas, para cinco das nove unidades mapeadas, com resultado superior para a classificação com o conjunto composto por 18 variáveis (CXbd1, LVd + LVA, RLd + RRd e RY). As demais quatro unidades não tiveram diferenças estatísticas entre os resultados obtidos pelos dois conjuntos (PVAd, PVd + PVAd, CXbd, LVd + PVd e AR + RLd).

O maior desempenho observado na classificação pelas redes com menor dimensão (18 variáveis) pode estar relacionado à capacidade do classificador em generalizar informações, entendido como fenômeno de Hughes, que segundo Hirose, et al. (1991), provoca um decréscimo na exatidão do classificador devido ao uso de variáveis redundantes ou de pouca significância. De acordo com Foody & Arora (1997), a exatidão de uma classificação aumenta com o uso de variáveis, até um ponto em que a adição de outras resulta em decréscimo e não apresenta efeito significativo.

O teste de significância revelou que, embora não tenham ocorrido diferenças significativas na classificação, quando se avaliou a exatidão dos mapas gerados pelo MaxVer e redes neurais, as unidades desses mapas foram discriminadas com diferentes índices de exatidão, diferindo entre uma e outra abordagem empregadas e em relação aos conjuntos de variáveis utilizadas. Assim, para algumas classes, a abordagem por redes neurais apresentou-se mais apropriada, com maior exatidão, enquanto que para outras, a exatidão não diferiu entre os dois métodos.

Os resultados corroboram com o exposto por Chagas (2005), sobre o efeito das características dos classificadores e das classes na exatidão do mapeamento. A identificação de diferenças estatísticas entre classificadores permite escolher aquele mais apropriado para determinado propósito, considerando-se a eficiência do classificador, o tempo gasto no processamento e as variáveis disponíveis para o mapeamento.

Quadro 21 – 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			Redes (18)	3,7			Redes (18)	8,7		
Redes (19)	0,2	8,2		Redes (19)	1,4	4,1		Redes (19)	10,9*	6,8	
MaxVer	5,5*	11,3*	7,8	MaxVer	4,4*	0,5	5,2	MaxVer	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			Redes (18)	8,7			Redes (18)	7,1		
Redes (19)	0,5	5,4		Redes (19)	0,5	8,1		Redes (19)	9,2*	5,9	
MaxVer	2,9*	7,0*	4,5	MaxVer	7,0*	0,5	9,4	MaxVer	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			Redes (18)	7,2			Redes (18)	9,4		
Redes (19)	9,6*	6,4		Redes (19)	9,6*	6,4		Redes (19)	11,0*	6,6	
MaxVer	5,3*	10,7*	12,0	MaxVer	1,4	7,3*	3,9	MaxVer	11,5*	2,4*	0,0

* Há diferença significativa a 5% de significância. Valor crítico de Z: 1,96.

Quadro 22 – 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	0,62	0,009267	8,2	0,59	0,007286	9,0	0,50	0,008147	7,8
PVd + PVAd	0,55	0,031891	4,1	0,32	0,023093	3,7	0,64	0,023794	5,2
CXbd1	0,21	0,004596	6,8	0,25	0,003296	8,7	0,28	0,002557	10,4
CXbd	0,67	0,022893	5,4	0,59	0,019833	5,4	0,37	0,018571	4,5
LVd + PVd	0,69	0,010537	8,1	0,64	0,008369	8,7	0,75	0,008461	9,4
LVd + LVA	0,33	0,009508	5,9	0,36	0,007088	7,1	0,18	0,006801	5,2
RLd + RRd	0,14	0,003335	6,6	0,50	0,003622	11,7	0,62	0,004286	12,0
RYd	0,76	0,018390	6,4	0,76	0,014725	7,2	0,15	0,009646	3,9
AR + RLd	0,73	0,016919	6,6	0,82	0,009244	9,4	1,00	0,000000	0,0
Valores obtidos para o mapa	kappa	Variância	Z	kappa	Variância	Z	kappa	Variância	Z
	0,51	0,00092	23,5	0,56	0,000898	25,1	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).

4.4.2. Validação do mapa gerado pelo método convencional

A concordância das classes de solo por unidades do mapa gerado pelo método convencional foi verificada com pontos de referência terrestre, calculando-se a exatidão para cada uma das classes de solos associadas às unidades. A exatidão global foi obtida pelo somatório dos pontos concordantes classificados em qualquer um dos componentes, considerando-se a classificação dos solos até o nível de subordens.

O método empregado na validação do mapa gerado por método convencional foi a estatística multivariada, por meio dos indicadores, índice kappa e exatidão global, assim como feito para os mapas classificados nas abordagens automáticas. As matrizes geradas para os solos do primeiro, segundo e terceiro componentes de associação das unidades são apresentadas respectivamente nos quadros 9, 10 e 11 (Anexo).

Os solos do primeiro componente, com índice kappa inferior a 0,40, apresentaram baixa concordância ($kappa = 0,38$), conforme os intervalos propostos por Landis e Koch (citados por Congalton e Green 1999) para avaliar o índice de acerto de uma classificação automática. O mesmo ocorreu para a concordância entre os solos do segundo e do terceiro componentes, que apresentaram índice kappa de 0,17 e 0,12, respectivamente para o primeiro e o segundo (Quadros 10 e 11 – Anexo).

Nos quadros 12, 13 e 14 (Anexo) são apresentadas as matrizes com a exatidão ponderada, respectivamente para os solos do primeiro, segundo e terceiro componentes das unidades de solo do mapa elaborado pelo método convencional. Os valores das matrizes de confusão foram convertidos para porcentagem de acertos usando a exatidão do produtor. Esse procedimento elimina o efeito do tamanho da amostra e permite a comparação direta entre as unidades (Congalton e Green, 1999).

Nas matrizes ponderadas a comparação dos erros de classificação são feitas de forma direta. Para os solos do primeiro componente (Quadro 12 – Anexo) devem ser ressaltadas as seguintes observações: os solos classificados como Cambissolos tiveram 74,4% das amostras confundidas com Latossolos Vermelho-Amarelos. Essa última classe também se confundiu com os Cambissolos em 20% dos pontos amostrados, e em 28%, com Argissolos Vermelhos. Os Latossolos Vermelhos tiveram confundimento em 50% dos pontos amostrados com os Latossolos Vermelho-Amarelos. Os Argissolos Vermelhos, que não tiveram nenhuma amostra corretamente classificada foram confundidos em 60,9% com os Argissolos Vermelho-Amarelos.

A exatidão global das classes de solos dos componentes das unidades do mapa é apresentada no quadro 23. Os solos do primeiro componente de associação apresentaram 47,4% de exatidão; 19,6% foi apresentado pelos solos do segundo, e 14,9% pelos solos do terceiro. O mapa apresentou exatidão global de 81,8%, obtida pela concordância de todos os pontos corretamente classificados em qualquer uma das classes de solos que compõem as associações das unidades.

Os solos do primeiro componente tiveram maior índice apresentado pelos Argissolos Vermelho-Amarelos, com 84% de acerto; já o menor índice ocorreu na classe dos Latossolos Vermelhos, com 19% dos pontos corretamente classificados. As demais classes de solos apresentaram entre 44 e 64% de exatidão. Considerando-se os acertos para os três componentes, o maior índice foi apresentado pelos Cambissolos (98,9%), seguido pelos Argissolos Vermelho-Amarelos, e Neossolos Litólicos, com 95,6 e 82% de exatidão, respectivamente. Valores entre 46,2 e 69% foram obtidos para os Neossolos Flúvicos, Argissolos Vermelhos e Latossolos Vermelhos (Quadro 23).

Quadro 23 – Concordância entre os pontos de referência de solos e os componentes de associação das unidades de mapeamento do mapa convencional

Classe	1º comp.	2º comp.	3º comp.	1º, 2º, 3º comp.
	Exatidão global (%)			
AR	55,0	-	0,0	55,0
CXbd	44,2	5,3	49,5	98,9
LVAd	52,0	-	28,0	80,0
LVd	18,5	50,0	-	68,5
PVAd	84,4	11,1	-	95,6
PVd	-	60,9	-	60,9
RLd	64,2	23,5	-	87,7
RY	46,2	-	-	46,2
RRd	-	-	0,0	0,0
Total	47,4	19,6	14,9	81,8

(^c) A classe de solos não ocorre como componente da unidade.

A concordância obtida para os solos do primeiro componente, em um mapa avaliado por Chagas (2006) em escala de 1:50.000, apresentou exatidão superior em 6%, em relação à obtida no presente estudo. Essa diferença indica que os resultados do presente trabalho podem ser considerados satisfatórios, já que o mapa utilizado apresenta escala duas vezes menor do que a do trabalho citado.

Os resultados da validação, por classes de solos, das unidades de mapeamento são apresentados no quadro 24, onde pode se observar que o menor índice refere-se à unidade CXbd1, com 25% dos pontos corretamente classificados. Já a unidade LVd1 apresentou 71,4% de acerto e concordância superior a 80% para as demais unidades.

A concordância dos solos para as unidades LVAd e CXbd3 mostra valores de exatidão decrescentes no sentido do primeiro para o terceiro componente. As unidades PVAd e LVd, por outro lado, apresentam exatidão superior para os solos do primeiro componente, conforme esperado, devido a ordem de importância da classe de solos dentro a unidade, tendo em vista que os solos do primeiro componente são mapeados com maior probabilidade de ocorrência. A unidade RLd apresentou 50% de acerto para cada um dos dois componentes da unidade (Quadro 24).

Quadro 24 – Valores de exatidão para os componentes de associação e unidades de mapeamento do mapa gerado pelo método convencional

Unidade de Mapeamento	1º comp.	2º comp.	3º comp.	Unidade
	Exatidão global (%)			
Ar1	AR 100,0	RLd 0		100,0
Ar3	AR 100,0	RLd -		100,0
CXbd	CXbd -	RLd 100,0	LVAd -	100,0
CXbd1	CXbd 100,0	LVA 0		25,0
CXbd2	CXbd 74,1	PVAd 7,4		81,5
CXbd3	CXbd 33,3	RLd 66,7		100,0
LVd	LVd 90,9	PVAd 9,1		100,0
LVAd	LVAd 14,0	LVd 23,7	CXbd 43,0	80,6
LVAd1	LVAd -	LVd 35,7	CXbd 35,7	71,4
PVAd	PVAd 54,3	PVd 11,4	LVAd 10,0	84,3
RLd	RLd 50,0	CXbd 50,0		100,0
RLd1	RLd 100,0	RRd -	AR -	100,0
RLd2	RLd 100,0	RRd -		100,0
RLd3	RLd 87,2	OOy -	CXbd 5,1	89,7
RLd4	RLd 88,9			88,9
RY	RY 100,0			100,0

(^o) Solos sem pontos amostrados - solos da unidade AR2, RLh e do segundo componente da unidade RLd3 (OOy), os quais não tiveram pontos de referência coletados para a validação do mapeamento. A exatidão global das unidades foi calculada pela razão entre o total de pontos amostrados e total de pontos corretamente classificados na unidade.

5. CONCLUSÕES

A avaliação da contribuição das variáveis para discriminar os solos em cada uma das abordagens empregadas foi uma importante etapa para seleção daquelas variáveis que permitiram os classificadores apresentarem maior exatidão no mapeamento. As variáveis que contribuíram em ordem crescente de importância para discriminar os solos na classificação pelas redes neurais foram: radiação solar, declividade, geologia e o índice *Iron oxide*. Na classificação pelo MaxVer, as variáveis categóricas, geologia, curvatura e aspecto não apresentaram contribuição para discriminar os solos, concorrendo com os maiores obtidos pelo classificador.

As avaliações dos parâmetros para a classificação pelas redes neurais, número de ciclos de treinamento e número de neurônios na camada interna das redes mostraram-se importantes para obter maior exatidão na classificação.

Os classificadores MaxVer e redes neurais apresentaram diferenças estatísticas significativas nos resultados do treinamento. No entanto, na validação com pontos de referência terrestre, os resultados não diferiram estatisticamente. Os mapas melhores classificados apresentaram índice kappa de 0,60 pelas redes neurais e de 0,58 pelo MaxVer.

Embora a exatidão global dos mapas tenha sido estatisticamente igual, os classificadores apresentaram diferenças na discriminação das unidades. A maioria das unidades que não apresentaram exatidão estatisticamente iguais nas duas abordagens foram melhores classificadas pelas redes neurais.

Os resultados obtidos nas abordagens testadas ressaltam a importância de se avaliarem, tanto a exatidão do mapa, quanto as suas categorias, observando-se a ocorrência de diferenças estatísticas em função do classificador e/ou conjuntos de variáveis discriminantes, e assim selecionar o mais apropriado para determinada classificação.

A classificação pelas redes neurais consumiu mais tempo do que pelo MaxVer, porém, os resultados foram superiores para a maioria das unidades de solos na primeira abordagem. A maior parte do tempo gasto pelas redes neurais ocorreu durante o treinamento do classificador, quando foram definidas as unidades de mapeamento, as variáveis discriminantes e a arquitetura das redes.

A concordância dos pontos de referência terrestre com os solos mapeados pelo método convencional foi de 85%. O mapa apresentou concordância de 48%, 19% e 14%, respectivamente para os solos do primeiro, segundo e terceiro componentes associados às unidades de solo.

REFERÊNCIAS BIBLIOGRÁFICAS

ATKINSON, P. M.; TATNALL, A. R. L. Neural networks in remote sensing. **International Journal of Remote Sensing**, 1997, v. 18, n. 4, p.699-709.

BACIC, I. L. Z. Demand-Driven Land Evaluation. In: Hartemink, A. E.; McBratney, A.; Mendonça-Santos M. de L. (Eds). **Digital soil mapping with limited data**. Springer: Netherlands, p.138-151. 2008.

BEHRENS, T. et al. An Approach to Removing Uncertainties in Nominal Environmental Covariates and Soil Class Maps. In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital soil mapping with limited data**. Springer: Netherlands, p.204-213. 2008.

BENEDIKTSSON, A. J.; SVEINSSON, J.R. Feature extraction for multisource data classification with artificial neural network, **International Journal of Remote Sensing**, v. 18, p.727-740. 1997.

BISHOP, et al. Uncertainty analysis for soil-terrain models. **International Journal of Geographical Information Science**. v. 20, n. 2, p.117-134. 2006.

BISHOP, T. F. A. et al. Uncertainty analysis for soil-terrain models, **International Journal of Geographical Information Sciences**, v.20, n.2, p.117-134. 2006.

BISHOP, T. F. A. et al. Measuring the quality of digital soil maps using information criteria. **Geoderma**. v.103, p. 95-111. 2001.

BOOTS, B.; CSILLAG, F. Categorical maps: comparisons, and confidence. **Journal of Geographical Systems**. v.8, p.109-118. 2006.

BUI, E. N. et al. Dokuchaev re-visited. In: **2nd Global Workshop on Digital Soil Mapping**. Rio de Janeiro: Embrapa Solos. p.17-26. 2006.

BURROUGH, P. A. et. al. The state of the art in pedometrics. **Geoderma**. v.62, p.311-326. 1994.

BURROUGH, P. A.; McDONNELL, R. A. **Principles of Geographical Information Systems**. New York: Oxford University Press Inc. 2004. 332p.

CARRÉ, F. et al. From a large to a small scale soil map: top-down against bottom-up approaches. In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital Soil Mapping with Limited Data**. Springer: Netherlands. p.194-203. 2008.

CARRÉ, F.; BOETTINGER, J.L. Synthesis and Priorities for FutureWork in Digital Soil Mapping. In: Hartemink, A. E.; McBratney, A.; Mendonça-Santos M. de L. (Eds). **Digital soil mapping with limited data**. Springer: Netherlands, p.382-399. 2008.

CARVALHO JUNIOR, W. de. **Classificação supervisionada de paisagens de áreas montanhosas utilizando redes neurais artificiais através de atributos geomorfométricos**. 2005. 143f. Tese (Doutorado em Ciências do solo). Universidade Federal de Viçosa, Viçosa: Minas Gerais. 2005.

CETEC – Centro de Tecnologia do Estado de Minas Gerais. **Mapa de solos do Estado de Minas Gerais**: Escala 1:600,000. 2008.

CHAGAS, C. da S. **Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros**. 2006. 223f. Tese (Doutorado em Ciências do Solo). Universidade Federal de Viçosa, Viçosa: Minas Gerais. 2006.

CHELLAIAH, S. Mapping of ultramafics and associated lithological units using satellite data. In: **2nd Asian Conference on GIS, GPS, Aerial Photography and Remote Sensing**. 3p. 2003.

COMIG - Companhia Mineradora de Minas Gerais, Projeto Espinhaço: **Geologia da Folha Baldim e Folha Conceição do Mato Dentro**, In: GROSSI-SAD, J, H, et al., (Org), CD-ROM, Belo Horizonte: COMIG. 1997.

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.

COOK, S. et al. Digital soil mapping for the 21st century: where to from here? A new global demand for digital soil information, In: **2nd Global Workshop on Digital Soil Mapping**, 2006, Rio de Janeiro: Embrapa Solos. 2006. 12p.

CPRM - Companhia de Pesquisa de Recursos Minerais, **Carta Geológica do Brasil ao Milionésimo, Folha SE, 23 - Belo Horizonte**, Serviço Geológico do Brasil, Ministério de Minas e Energia, Secretária de Minas e Metalurgia, Escala 1:1000,000, 2004, CD-ROM.

DEMATTE, J. A. M. et al. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. **International Journal of Remote Sensing**, v.28, 2007. p. 3813-3829.

DOBOS, E. et al. Use of combined digital elevation model and satellite data for regional soil mapping, **Geoderma**. v.97, p.367-391. 2000.

ESDI. Earth Science Data Interface. Imagem LANDSAT 7. Disponível em:< <http://glcfapp.umiacs.umd.edu:8080/esdi/index.jsp>> Acesso em 20 de Jun. de 2006.

EMBRAPA/SPI. Empresa Brasileira de Pesquisa Agropecuária, Sistema de Produção de Informação. **Sistema Brasileiro de Classificação de Solos**. Rio de Janeiro: Embrapa SPI. 2006. 306p.

EMBRAPA/CNPS. Empresa Brasileira de Pesquisa Agropecuária. Centro Nacional de Pesquisa de Solos. **Procedimentos Normativos de Levantamentos Pedológicos**. Brasília. 1995. 111p.

EMBRAPA. Serviço Nacional de Levantamento e Conservação de Solos. **Manual de métodos de análise de solo**. Rio de Janeiro. v1. 1979.

Environmental System Research Institute (ESRI), ArcGIS 9.2. New York, 2006.

- FIDÊNCIO, P. H. et al. Application of artificial neural networks to the classification of soils from São Paulo state using near-infrared spectroscopy. **Analyst**. v.126, p.2194-2200. 2001.
- FOODY, G. M.; ARORA, M.K. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. **International Journal of Remote Sensing**. v.18, n.4, p.799-810. 1997.
- FOODY, G. M. What is the difference between two maps? A remote sensor's view. **Journal of Geographical Sys**. v. 8, n.2, p.119-130. 2006.
- FOODY, G. M; ARORA, M. K. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. **International Journal of Remote Sensing**. v.18, n.4, p.799-810. 1997.
- FUNDAÇÃO CETEC, Solos, In: **Diagnóstico ambiental do Estado de Minas Gerais**, Belo Horizonte. p.31-34 (Série de Publicações Técnicas). 1983.
- GESSLER, P. E. et al. Modeling Soil–Landscape and Ecosystem Properties Using Terrain Attributes. **Soil Science Society of American Journal**. v.64. p.2046–2056. 2000.
- HAYKIN, S. **Neural Networks: a comprehensive foundation**. Delhi: Pearson. 2nd Ed. 2001. 823 p.
- HENGL, T. **Pedometric mapping: bridging the gaps between conventional and pedometric approaches**. 2003. 200f. (PhD thesis) - Wageningen University, Enschede, The Netherlands. 2003.
- 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.
- HIROSE, Y., et al. 1991, Back-propagation algorithm which varies the number of hidden units. **Neural Networks**, v. 4, p.61- 66.
- HOWELL, D. et al. Development and Application of Digital Soil Mapping Within Traditional Soil Survey: what will it grow into? In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital Soil Mapping with Limited Data**. Springer: Netherlands. p.32-43. 2008.
- HUDSON, B. D. The soil survey as a paradigm-based science. **Soil Science Society of American Journal**. v.56, p.836-841.1992.
- IBAMA – Instituto Nacional do Meio Ambiente e dos Recursos Naturais Renováveis. Disponível em<<http://www.ibama.gov.br/siucweb/mostraUc.php?seqUc=3>> Acesso em: 20 setembro de 2006.
- IBAMA. Instituto Nacional do Meio Ambiente e dos Recursos Naturais Renováveis, **Mapa das Unidades de conservação Federal do Brasil**. Escala 1: 6.000.000. 2006a.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Manual técnico de pedologia**. 2 ed. n. 4, Rio de Janeiro: IBGE. 2007. 316p, (Manuais técnicos em geociências).

- IBGE - Instituto Brasileiro de Geografia e Estatística. Projeção Universal Transversa de Mercator. **Carta SE-23-Z-C-VI-2 / MI 2535-2 - Jaboticatubas**, Escala 1:50.000. 1977.
- IBGE - Instituto Brasileiro de Geografia e Estatística. Projeção Universal Transversa de Mercator. **Carta SE-23-Z-C-III/MI 2497 – Baldim**. Escala 1:100.000. 1977.
- IBGE - Instituto Brasileiro de Geografia e Estatística. Projeção Universal Transversa de Mercator. **Carta SE-23_Z-D-I /MI 2498 - Conceição do Mato Dentro**. Escala 1:100.000. 1977.
- IBGE - Instituto Brasileiro de Geografia e Estatística. Projeção Universal Transversa de Mercator. **Carta SE-23-Z-D-IV /MI 2536 – Itabira**. Escala 1:100.000. 1977.
- IEF/MG - Instituto Estadual de Florestas de Minas Gerais. Disponível em: <<http://www.ief.mg.gov.br/parques/areasprot.asp>>. Acesso em 16 de Setembro de 2006.
- INDORANTE, S. J. et al. Positioning soil survey for the 21st century. **Journal of Soil and Water Conservation**. v.51, p.21-28, 1996.
- JAVANNS. Java Neural Network Simulator Version 1.1. University of Tübingen. Disponível em: <http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/welcome_e.html>. Acesso em 4 março de 2008.
- JENNY, H. **Factors of soil formation**: a system of quantitative pedology. New York: Mcgraw. 1941. 281p.
- LILLESAND, T. M.; KIEFER, R. W. **Remote sensing and image interpretation**. John Wiley & Sons. New York, 3rd ed, 1994.
- KAVZOGLU, T.; MATHER P. M. Pruning artificial neural networks: an example using land cover classification of multi-sensor images. **International Journal of Remote Sensing**. 1999. v. 20, p. 2787-2830.
- KAVZOGLU, T.; MATHER, P.M. The use of backpropagating artificial neural network in land cover classification. **International Journal of Remote Sensing**. v.24, p.4907-4938, 2003.
- KROL, B. G. C. M. Towards a Data Quality Management Framework for Digital Soil Mapping with Limited Data. In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital Soil Mapping with Limited Data**. Springer: Netherlands. 2008. p.114-137.
- KANELLOPOULOS, I.; WILKINSON, G. G. Strategies and best practice for neural network image classification. **International Journal of Remote Sensing**, 1997, v. 18, n. 4, p. 711- 725.
- KROL, B. G. C. M. Towards a soil data quality management framework for predictive soil mapping at intermediate scales in data-poor environments. In: **2nd Global Workshop on Digital Soil Mapping**. Rio de Janeiro: Embrapa Solos. 2006. 7p.
- LAGACHERIE, P. Digital Soil Mapping: A state of the art. In: **2nd Global Workshop on Digital Soil Mapping**. Rio de Janeiro: Embrapa Solos. 2006. 9p.

- MATHER, P. M. **Computer Processing of Remotely-Sensed Images: An Introduction**, 3 ed. Chichester: John Wiley & Sons. 2004. 350p.
- McBRATNEY, A. B. et al, An overview of pedometric techniques for use in soil survey. **Geoderma**. v.97, p.293-327. 2000.
- McBRATNEY, A. B. et al. On digital soil mapping. **Geoderma**. v.117, p.3-52. 2003.
- McBRATNEY, A. B. et al. An overview of pedometric techniques for use in soil survey. **Geoderma**. v.97, p.293-327. 2000.
- McKENZIE, N. J. & RYAN, P. J. Spatial prediction of soil properties using environmental correlation. **Geoderma**. v.89, p.67-94. 1999.
- MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto: metodologias de aplicação**. 3 ed. Viçosa: UFV. 2005. 320p.
- MORAN, C. J.; BUI, E. N. Spatial data mining for enhanced soil map modelling. **International Journal of Geographical Information Science**. v.16, p.533-550. 2002.
- MOORE, I. D. Soil attribute predicting using terrain analysis. **J Soil Sci. Soc. Am.** n. 57, p.443-453. 1993.
- PURNELL, M. F. Soil survey information supply and demand: international policies and stimulation programmes. In: ZINCK, J. A. **Soil survey: perspectives and strategies for the 21st century**. ITC Publication. Enschede. The Netherlands. n. 21, p.30-35. 1993.
- RAJESH, H.M. Application of remote sensing and GIS in mineral resource mapping: an overview. **Journal of Mineralogical and Petrological Sciences**. v.99, p. 83-103, 2004.
- RIBEIRO, S. R. A. **Integração de imagens Landsat TM e dados auxiliares para a delimitação automática de unidades fisionômicas (paisagem) e geológicas usando redes neurais artificiais**. 2003. 169f. Tese (Doutorado em Ciências Geodésicas). Universidade Federal do Paraná, Curitiba: Paraná, 2003.
- ROSSITER, D. G. Assessing the thematic accuracy of area-class soil maps. **Soil Science Division**. International Institute for Aerospace Survey and Earth Sciences. ITC Publication. 2001. 46p.
- ROSSITER, D. G. Digital soil mapping: towards a multiple-use Soil Information System. **Análisis Geográficos**. Augustin Codazzi. n.32, v.1. 2005.
- ROSSITER, D. G. Digital soil resource inventories: status and prospects. **Journal of Soil Use and Management**. n. 20, v. 3, p.296-301. 2004.
- ROSSITER, D. G. **Methodology for soil resource inventories**: 2 ed. ITC Publication. 2000. 132p.
- 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.

SCULL, P. et al, 2003. **The application of classification tree analysis to soil type prediction in a desert landscape.** Ecological Modelling. n.181, 2005. p.1–15.

SIRTOLI, A. E. **Mapeamento de solos com auxílio da geologia, atributos do terreno e índices espectrais integrados por redes neurais artificiais.** 2008, 114f. Tese (Doutorado em geologia). Universidade Federal do Paraná, Curitiba: Paraná, 2008.

SNUC. **Sistema Nacional de Unidades de Conservação.** Lei N° 9.985. Brasília: Distrito Federal. 2000. 20p.

THOMAS, A. L. et al. A. Predicting soil classes with parameters derived from relief and geologic materials in a sandstone region of the Vosges mountains (Northeastern France). *Geoderma*, v.90, p.291-305. 1999.

VINCENT, R.K. **Fundamentals of Geological and Environmental Remote Sensing.** Prentice Hall. 1997. New Jersey, 366 pp.

VIEIRA, C. A. O. **Accuracy of remotely sensing classification of agricultural crops: a comparative study.** 2000. 327f. (PhD Thesis). University of Nottingham, UK. 2000.

ZELL, A. et al. *SNNS (Stuttgart Neural Network Simulator).* User Manual v.4.2. Disponível em: <<http://www.ra.cs.uni-tuebingen.de/SNNS/>> Acesso em 14 de jan de 2008.

ZHU, A. X. et al. Purposive Sampling for Digital Soil Mapping for Areas with Limited Data. In: Hartemink, A. E.; McBratney, A. & Mendonça-Santos M. de L. (Eds). **Digital Soil Mapping with Limited Data.** Springer: Netherlands. p.226-233. 2008.

ZHU, A. X. et al. Soil mapping using GIS, expert knowledge, and fuzzy logic. **Soil Science Society of American Journal.** v.65, p.1463-1472. 2001.

ZHU, A. X. Mapping soil landscape as spatial continua: the neural network approach. **Water Resources Research,** v.36, p.663-677. 2000.

ZHU, A. X. A similarity model for representing soil spatial information. **Geoderma.** v.77, p.217-242. 1997.

ANEXOS

Quadro 1 – Matriz de confusão para a classificação feita pelas redes neurais sem a variável NDVI

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão Usuário	Z
PVAd	188	1	0	0	10	0	0	0	1	200	94	49.712
PVd + PVAd	0	194	0	0	1	7	0	0	0	202	96	62.186
CXbd1	0	0	197	4	0	2	1	0	0	204	96.6	67.33
CXbd	1	0	0	195	14	0	0	0	2	212	92	43.795
LVd + PVd	6	0	0	0	165	2	3	0	0	176	93.8	45.587
LVd + LVA	0	5	2	0	0	189	1	0	0	197	95.9	60.606
RLd + RRd	3	0	1	0	6	0	194	1	0	205	94.6	53.416
RYd	0	0	0	0	4	0	0	199	1	204	97.5	80.08
AR + RLd	2	0	0	1	0	0	1	0	196	200	98	87.979
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	94	97	98.5	97.5	82.5	94.5	97	99.5	98			

Kappa = 0,948; Exatidão global = 95,4; Variância = 0,000031; Z estatístico = 170.536, Z Tab = 1,96.

Quadro 2 – Matriz de confusão para a classificação feita pelas redes neurais para o conjunto sem a variável aspecto

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão usuário	Z
PVAd	191	1	1	0	5	7	2	0	0	207	92.3	44.141
PVd + PVAd	0	190	0	0	0	3	0	0	0	193	98.4	98.255
CXbd1	0	0	199	1	0	0	0	0	0	200	99.5	177.326
CXbd	1	0	0	197	7	0	2	0	4	211	93.4	48.405
LVd + PVd	8	0	0	1	175	5	1	0	0	190	92.1	41.771
LVd + LVA	0	9	0	0	0	185	2	0	0	196	94.4	50.986
RLd + RRd	0	0	0	1	9	0	193	5	0	208	92.8	45.945
RYd	0	0	0	0	4	0	0	195	0	199	98	87.533
AR + RLd	0	0	0	0	0	0	0	0	196	196	100	0
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	95,5	95	99,5	98,5	87,5	92,5	96,5	97,5	98			

Kappa = 0,951; Variância = 0,000029; Z estatístico = 175,025; Z Tabelado = 1,96.

Quadro 3 – Matriz de confusão do treinamento da classificação feita pelo algoritmo MaxVer com 19 variáveis discriminantes

Unidades	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidã Usuário	Z
PVAd	0	0	0	0	0	0	0	0	0	0	0	0
PVd + PVAd	0	0	0	0	0	0	0	0	0	0	0	0
CXbd1	0	0	197	0	0	2	0	0	0	199	99	124,438
CXbd	0	0	0	0	0	0	0	0	0	0	0	0
LVd + PVd	189	116	0	181	155	0	1	0	60	702	22,1	10,027
LVd + LVA	7	76	0	12	0	198	0	133	0	426	46,5	16,911
RLd + RRd	4	8	3	7	45	0	199	67	140	473	42,1	16,271
RYd	0	0	0	0	0	0	0	0	0	0	0	0
AR + RLd	0	0	0	0	0	0	0	0	0	0	0	0
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	0	0	98,5	0	77,5	99	99,5	0	0			

Kappa = 0,34; Exatidão global = 41,6; Variância = 0,000126; Z estatístico = 30,511; Z Tab = 1,96.

Quadro 4 – Matriz de confusão obtida no treinamento da classificação pelo MaxVer sem a variável geologia

Unidade	PVAd	PVd + PVAd	CXbd1	CXbd	LVd + PVd	LVd + LVA	RLd + RRd	RYd	AR + RLd	Total	Exatidão Usuário	Z
PVAd	181	0	0	0	16	0	0	5	0	202	89,6	37,01
PVd + PVAd	1	185	0	0	0	38	0	6	0	230	80,4	27,225
CXbd1	0	2	197	0	1	2	0	0	0	202	97,5	79,282
CXbd	0	0	0	199	0	0	6	0	0	205	97,1	73,278
LVd + PVd	15	2	1	0	152	0	0	3	1	174	87,4	30,66
LVd + LVA	2	11	0	0	0	160	0	138	0	311	51,4	15,607
RLd + RRd	1	0	2	0	30	0	194	47	1	275	70,5	22,707
RYd	0	0	0	0	0	0	0	0	0	0	0	0
AR + RLd	0	0	0	1	1	0	0	1	198	201	98,5	102,368
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	90,5	92,5	98,5	99,5	76	80	97	0	99			

Kappa = 0,791; Exatidão global = 81,4; Variância = 0,000103; Z estatístico = 77,936; Z Tab = 1,96.

Quadro 5 – Matriz de confusão da classificação pelo MaxVer sem as variáveis discriminantes geologia e curvatura

Unidade	PVAd	PVd + PVAd	CXbd1	CXbd	LVd + PVd	LVd + LVA	RLd + RRd	RYd	AR + RLd	Total	Exatidão Usuário	Z
PVAd	180	7	0	0	18	0	0	8	0	213	84,5	30,191
PVd + PVAd	2	179	0	0	2	38	0	0	0	221	81	27,152
CXbd1	0	2	195	0	1	2	0	0	0	200	97,5	78,485
CXbd	0	0	0	199	0	0	5	0	0	204	97,5	80,08
LVd + PVd	16	4	2	0	147	0	1	0	1	171	86	28,584
LVd + LVA	2	8	0	0	0	160	0	51	0	221	72,4	21,139
RLd + RRd	0	0	3	0	31	0	194	7	1	236	82,2	29,28
RYd	0	0	0	0	0	0	0	134	0	134	100	0
AR + RLd	0	0	0	1	1	0	0	0	198	200	99	125,06
Total	200	200	200	200	200	200	200	200	200	1800		
Exat. Prod.	90	89,5	97,5	99,5	73,5	80	97	67	99			

Kappa = 0,866; Exatidão global = 88,1; Variância = 0,000073; Z estatístico = 101,097; Z Tab = 1,96; Variáveis utilizadas: 17 (sem geologia e curvatura).

Quadro 6 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelas redes neurais com todas as 19 variáveis

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão Usuário	Z
PVAd	26	0	1	0	7	0	1	0	0	35	74,3	8.166
PVd + PVAd	0	9	9	0	1	16	3	0	0	38	23,7	4.136
CXbd1	0	1	36	0	0	6	15	0	0	58	62,1	6.834
CXbd	0	0	0	7	0	0	6	0	1	14	50,0	5.397
LVd + PVd	12	0	0	1	26	1	8	0	0	48	54,2	8.107
LVd + LVA	0	4	29	0	0	35	1	3	0	72	48,6	5.895
RLd + RRd	0	0	11	2	0	1	29	0	3	46	63,0	6.559
RYd	0	0	0	0	0	0	0	11	0	11	100,0	6.410
AR + RLd	0	0	0	1	0	0	24	0	15	40	37,5	6.586
Total	38	14	86	11	34	59	87	14	19	362		
Exat. Prod.	68,4	64,3	41,9	63,6	76,5	59,3	33,3	78,6	78,9			

Kappa = 0,531; Exatidão global = 54,3; Variância = 0,00092; Z calculado = 16,748; Z tab. 1,96.

Quadro 7 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelas redes neurais para o conjunto com 18 variáveis (sem a variável *Clay mineral*)

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidã Usuário	Z
PVAd	25	0	1	0	8	1	1	0	0	36	69,4	9.032
PVd + PVAd	2	6	10	0	0	14	2	0	0	34	17,6	3.715
CXbd1	0	2	37	0	0	6	14	0	0	59	62,7	8.744
CXbd	0	0	0	6	2	0	4	0	1	13	46,2	5.449
LVd + PVd	11	0	0	1	24	0	3	0	0	39	61,5	8.727
LVd + LVA	0	6	30	0	0	37	1	3	0	77	48,1	7.141
RLd + RRd	0	0	2	4	0	0	55	0	2	63	87,3	11.695
RYd	0	0	0	0	0	0	0	11	0	11	100,0	7.169
AR + RLd	0	0	6	0	0	1	7	0	16	30	53,3	9.394
Total	38	14	86	11	34	59	87	14	19	362		
Exat. Prod.	65,8	42,9	43,0	54,5	70,6	62,7	63,2	78,6	84,2			

Kappa = 0,56. Exatidão global = 60,1; Variância = 0,000898; Z calculado = 18,821; Z tab. 1,96.

Quadro 8 – Matriz de confusão obtida na validação com pontos de referência terrestre para a classificação feita pelo MaxVer para o conjunto com 11 variáveis (sem as variáveis geologia, curvatura e bandas espectrais da imagem de satélite)

Unidade	PVAd	PVd+ PVAd	CXbd1	CXbd	LVd+ PVd	LVd+ LVA	RLd+ RRd	RYd	AR+ RLd	Total	Exatidão Usuário	Z
PVAd	22	0	1	0	5	4	1	1	0	34	65	7.822
PVd + PVAd	0	10	9	0	1	19	2	1	0	42	24	5.175
CXbd1	0	1	35	0	0	4	1	0	0	41	85	10.450
CXbd	0	0	0	2	0	0	1	0	0	3	67	4.462
LVd + PVd	16	0	8	2	28	2	3	0	0	59	47	9.419
LVd + LVA	0	3	26	1	0	29	3	10	0	72	40	5.176
RLd + RRd	0	0	6	6	0	1	65	0	0	78	83	12.020
RYd	0	0	0	0	0	0	0	2	0	2	100	3.888
AR + RLd	0	0	1	0	0	0	11	0	19	31	61	1
Total	38	14	86	11	34	59	87	14	19	362		
Exat. Prod.	57,9	71,4	40,7	18,2	82,4	49,2	74,7	14,3	100,0			

Kappa = 0,547. Exatidão global = 59,0; Variância = 0,000869; Z calculado = 18,556; Z tab. 1,96.

Quadro 9 – Matriz de confusão para as classes de solo do Primeiro Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd	Exat. Usu.
Ar	11	0	0	0	0	0	5	0	0	69
CXbd	0	42	4	5	3	4	15	0	0	58
LVd	0	0	10	0	5	1	0	0	0	63
LVAd	0	45	27	13	0	2	8	7	0	13
PVd	0	0	0	0	0	0	0	0	0	0
PVAd	0	0	10	7	14	38	1	0	0	54
RLd	9	8	3	0	1	0	52	0	6	66
RY	0	0	0	0	0	0	0	6	0	100
RRd	0	0	0	0	0	0	0	0	0	69
Total	20	95	54	25	23	45	81	13	6	362
Exat. Prod.	55	44	19	52	0	84	64	46	0	

Kappa 0,379; Exatidão global 47,5; Variância 0,00091211.

Quadro 10 – Matriz de confusão para as classes de solo do Segundo Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd	Exat. Usu.
Ar	0	0	0	0	0	0	0	0	0	0
CXbd	9	5	2	0	1	0	5	0	3	21
LVd	0	45	27	13	5	2	8	5	0	48
LVAd	0	1	0	0	3	0	0	0	0	0
PVd	0	0	10	7	14	38	1	0	0	87
PVAd	0	40	14	5	0	5	1	0	0	17
RLd	11	1	0	0	0	0	19	0	0	76
RY	0	0	0	0	0	0	0	0	0	0
RRd	0	0	0	0	0	0	5	0	0	0
* Não clas.	0	3	1	0	0	0	42	8	3	
Total	20	95	54	25	23	45	81	13	6	362
Exat. Prod.	0	5	50	0	61	11	23	0	0	

Kappa 0,168; Exatidão global 19,6; Variância 0,000555; * Não classificada.

Quadro 11 – Matriz de confusão para as classes de solo do Terceiro Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd	Exat. Usu.
Ar	0	0	0	0	0	0	3	0	0	0
CXbd	6	47	28	13	5	2	47	7	3	30
LVd	0	0	0	0	0	0	0	0	0	0
LVAd	0	0	10	7	14	38	13	0	0	8
PVd	0	0	0	0	0	0	0	0	0	0
PVAd	0	0	0	0	0	0	0	0	0	0
RLd	0	0	0	0	0	0	0	0	0	0
RY	0	0	0	0	0	0	0	0	0	0
RRd	0	0	0	0	0	0	0	0	0	0
* Não clas.	14	48	16	5	4	5	18	6	3	0
Total	20	95	54	25	23	45	81	13	6	362
Exat. Prod.	0	49	0	28	0	0	0	0	0	

Kappa 0,117; Exatidão global 14,9; Variância 0,000757921; * Não classificada.

Quadro 12 – Matriz de confusão ponderada para as classes de solo do Primeiro Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd
Ar	55,0	0	0	0	0	0	6,2	0	0
CXbd	0	44,2	7,4	20	13,0	8,9	18,5	0	0
LVd	0	0	18,5	0	21,7	2,2	0	0	0
LVAd	0	47,4	50	52,0	0	4,4	9,9	53,8	0
PVd	0	0	0	0	0	0	0	0	0
PVAd	0	0	18,5	28,0	60,9	84,4	1,2	0	0
RLd	45,0	8,4	5,6	0	4,3	0	64,2	0	100
RY	0	0	0	0	0	0	0	46,2	0
RRd	0	0	0	0	0	0	0	0	0

Quadro 13 – Matriz de confusão ponderada para as classes de solo do Segundo Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd
Ar	0	0	0	0	0	0	0	0	0
CXbd	45,0	5,3	3,7	0	4,3	0	6,2	0	50
LVd	0	47,4	50	52,0	21,7	4,4	9,9	38,5	0
LVAd	0	1,1	0	0	13,0	0	0	0	0
PVd	0	0	18,5	28,0	60,9	84,4	1,2	0	0
PVAd	0	42,1	25,9	20	0	11,1	1,2	0	0
RLd	55,0	1,1	0	0	0	0	23,5	0	0
RY	0	0	0	0	0	0	0	0	0
RRd	0	0	0	0	0	0	6,2	0	0

Quadro 14 – Matriz de confusão ponderada para as classes de solo do Terceiro Componente de associação das unidades do mapa gerado pelo método convencional

Classe	Ar	CXbd	LVd	LVAd	PVd	PVAd	RLd	RY	RRd
Ar	0	0	0	0	0	0	3,7	0	0
CXbd	30	49,5	51,9	52,0	21,7	4,3	58,0	53,8	50
LVd	0	0	0	0	0	0	0	0	0
LVAd	0	0	18,5	28,0	60,9	85,1	16,0	0	0
PVd	0	0	0	0	0	0	0	0	0
PVAd	0	0	0	0	0	0	0	0	0
RLd	0	0	0	0	0	0	0	0	0
RY	0	0	0	0	0	0	0	0	0
RRd	0	0	0	0	0	0	0	0	0