

Comissão 1.3 - Pedometria

MAPEAMENTO PEDOLÓGICO DIGITAL DA FOLHA BOTUCATU (SF-22-Z-B-VI-3): TREINAMENTO DE DADOS EM MAPA TRADICIONAL E VALIDAÇÃO DE CAMPO⁽¹⁾

Cristiano Cassiano da Silva⁽²⁾, Ricardo Marques Coelho⁽³⁾, Stanley Robson de Medeiros
Oliveira⁽⁴⁾ & Samuel Fernando Adami⁽⁵⁾

RESUMO

O mapeamento digital de solos permite prever padrões de ocorrência de solos com base em áreas de referência e no uso de técnicas de mineração de dados para modelar associações solo-paisagem. Os objetivos deste trabalho foram produzir um mapa pedológico digital por meio de técnicas de mineração de dados aplicadas a variáveis geomorfométricas e de geologia, com base em áreas de referência; e testar a confiabilidade desse mapa por meio de validação em campo com diferentes sistemas de amostragem. O mapeamento foi realizado na folha Botucatu (SF-22-Z-B-VI-3), utilizando-se as folhas 1:50.000, Dois Córregos e São Pedro, como áreas de referência. Variáveis descritoras do relevo e de geologia associadas às unidades de mapeamento pedológico das áreas de referência compuseram a matriz de dados de treinamento. A matriz foi analisada pelo algoritmo PART de árvore de decisão, do aplicativo Weka (Waikato Environment for Knowledge Analysis), que cria regras de classificação. Essas regras foram aplicadas aos dados geomorfométricos e geológicos da folha Botucatu, para predição de unidades de mapeamento pedológico. A validação de campo dos mapas digitais deu-se por meio de amostragem por transectos em uma unidade de mapeamento da folha São Pedro e de forma aleatório-estratificada na folha Botucatu. A avaliação da unidade de mapeamento na folha São Pedro verificou confiabilidade, respectivamente, de 83 e 66 %, para os mapas pedológicos digital e tradicional com legenda simplificada. Apesar de terem sido geradas regras para todas as unidades de mapeamento pedológico das áreas de treinamento, nem todas as unidades de mapeamento foram preditas na folha Botucatu, o que resultou das diferenças de relevo e geologia

⁽¹⁾ Parte da Dissertação de Mestrado do primeiro autor apresentada ao programa de Agricultura Tropical e Subtropical do Instituto Agronômico, Campinas, SP. Com bolsa da Capes. Recebido para publicação em 20 de junho de 2012 e aprovado em 26 de abril de 2013.

⁽²⁾ Engenheiro Agrônomo, Mestre em Agricultura Tropical e Subtropical. E-mail: ccsilva2@yahoo.com.br

⁽³⁾ Pesquisador Científico, Instituto Agronômico. Av. Barão de Itapura, 1481. Caixa Postal 28. CEP 13012-970 Campinas (SP). E-mail: rmcoelho@iac.sp.gov.br

⁽⁴⁾ Pesquisador, Embrapa Informática Agropecuária. Av. André Tosello, 209. CEP 13083-886 Campinas (SP). E-mail: stanley.oliveira@embrapa.br

⁽⁵⁾ Professor Adjunto, Universidade Federal da Integração Latino-Americana. Av. Tancredo Neves, 6731, Bl. 4. Caixa Postal 2044. CEP 85867-970 Foz do Iguaçu (PR). E-mail: samuel.adami@unila.edu.br

entre as áreas de treinamento e de mapeamento. A validação de campo do mapa digital da folha Botucatu verificou exatidão global de 52 %, compatível com levantamentos em nível de reconhecimento de baixa intensidade, e kappa de 0,41, indicando qualidade Boa. Unidades de mapeamento mais extensas geraram mais regras, resultando melhor reprodução dos padrões solo-relevo na área a ser mapeada. A validação por transectos na folha São Pedro indicou compatibilidade do mapa digital com o nível de reconhecimento de alta intensidade e compatibilidade do mapa tradicional, após simplificação de sua legenda, com o nível de reconhecimento de baixa intensidade. O treinamento do algoritmo em mapas e não em observações pontuais reduziu em 14 % a exatidão do mapa pedológico digital da folha Botucatu. A amostragem aleatório-estratificada pelo hipercubo latino é apropriada a mapeamentos com extensa base de dados, o que permite avaliar o mapa como um todo, tornando os trabalhos de campo mais eficientes. A amostragem em transectos é compatível com a avaliação da pureza de unidades de mapeamento individualmente, não necessitando de base de dados detalhada e permitindo estudos de associações solo-paisagem em pedossequências.

Termos de indexação: modelagem de associações solo-paisagem, mineração de dados, qualidade de mapas pedológicos, hipercubo latino, amostragem em transectos, mapas de reconhecimento de solos.

SUMMARY: *DIGITAL PEDOLOGICAL MAPPING OF BOTUCATU SHEET (SF-22-Z-B-VI-3): DATA TRAINING ON CONVENTIONAL MAPS AND FIELD VALIDATION*

Digital soil mapping allows predicting patterns of soil classes on the basis of well-known reference areas and of data mining techniques to model soil-landscape relationships. The purpose of this study was to (1) generate a digital pedological map using data mining techniques to associate geomorphometric and geology variables with soil classes of traditional soil maps in reference areas and (2) validate these maps by different field techniques. The mapping was carried out using the 1:50.000 Botucatu sheet (SF-22-Z-B-VI-3), and 1:50.000 Dois Córregos and São Pedro sheets (São Paulo, Brazil) as reference areas. Training data - soil mapping units (MU) and topographic and geological variables from the reference areas were analyzed by PART, a decision-tree algorithm found on the Weka (Waikato Environment for Knowledge Analysis) software, producing classification rules, which were applied to the Botucatu sheet. Field validation of the produced digital maps was carried out by transect sampling in the zone of São Pedro and by a stratified-random sampling procedure at Botucatu sheet. Accuracy of the mapping unit at São Pedro was 83 %, for the digital and 66 %, for the traditional soil map with simplified legend. Although analysis generated rules for all MU's of the training areas, not all MU's were predicted on the Botucatu sheet, due to differences in relief and geology between training and mapping areas. Digital soil map of Botucatu had an overall accuracy of 52 %, consistent with reconnaissance soil surveys of low intensity, and a kappa index of 0.41, indicating good quality. Larger mapping units on training areas produced more rules, thus reproducing more accurately soil-landscape pattern of the mapped area. Validation at the São Pedro sheet by transect sampling suggested that the digital map is consistent to high intensity reconnaissance soil surveys; whereas the traditional map (simplified legend) quality corresponded to that of low-intensity soil surveys. Training of the algorithm on maps, not in field-observed points, reduced accuracy of digital soil map by 14 %. Latin hypercube sampling is adequate for mapping areas with large data bases, allows to evaluate the entire mapped area and imparts efficiency to field work. Transect sampling is adequate to evaluate purity of individual mapping units, requires no detailed data base of predictor variables, and allows studies on soil-landscape relationships in pedosequences.

Index terms: soil-landscape relationship modeling, data mining, quality of soil class maps, latin hypercube sampling, transect sampling, reconnaissance soil surveys.

INTRODUÇÃO

Tradicionalmente, no mapeamento pedológico é realizada uma avaliação prévia da área a ser mapeada por fotointerpretação ou por outras técnicas que permitem avaliar a sua fisiografia, com auxílio de outros dados do meio físico-biótico disponíveis, com intenso apoio de atividades de caracterização e coleta de solos em campo e de dados analíticos de laboratório. Os mapeamentos tradicionais têm pequeno apoio de técnicas digitais, utilizadas, atualmente, especialmente na sua representação cartográfica. O mapeamento digital de solos introduz algumas vantagens em relação ao método tradicional, apresentando-se como alternativa rápida e econômica (McBratney et al., 2003). Esse tipo de mapeamento pode ser definido como a criação de sistemas espaciais de informação, utilizando modelos numéricos para a inferência das variações espaciais dos tipos de solos, a partir de observações e conhecimento desses e de variáveis ambientais correlacionadas, como as variáveis geomorfométricas, declividade e curvaturas (Moore et al., 1993).

Uma das vantagens do mapeamento digital com base no conhecimento dos padrões regionais de solos é a possibilidade de prever a ocorrência de tipos de solos em áreas não mapeadas, com uso de informações geradas previamente em áreas de referência (Lagacherie & Voltz, 2000). A informação das áreas de referência pode ser obtida por técnicas de mineração de dados, que encontram padrões e geram conhecimento a partir de conjuntos de dados (Han & Kamber, 2001). Dentre as várias técnicas de mineração de dados, está a árvore de decisão, técnica de aprendizagem de máquina, que classifica e prediz amostras desconhecidas com base em registros conhecidos.

Para que um mapa digital produzido sem apoio de campo tenha confiabilidade, sua validação em campo é necessária, identificando-se e coletando-se o solo de forma a representar suas unidades de mapeamento. Por permitirem alguma sistematização das observações, a exemplo da densidade de amostragem, e assim poderem ser associados à escala do mapeamento, transectos de pontos de amostragem, também usados na execução de levantamentos de solos tradicionais, podem ser aplicados para validação dos mapeamentos (Silva, 2000; Bertolani, 2003). Transectos aleatórios podem ser utilizados onde a direção do gradiente de solos é desconhecida, enquanto transectos sistemáticos podem ser aplicados quando há um gradiente de solos esperado (Wilding, 1985). Menos tradicional nos levantamentos de solos é a amostragem estratificada pelo hipercubo latino. Na seleção de pontos de observação e de amostragem de solo pelo hipercubo latino, o número de sorteios de cada classe ou combinação de classes das variáveis selecionadas é diretamente proporcional à sua probabilidade de ocorrência (Minasny & McBratney,

2002), o que se adequa bem à estratificação da amostragem pela morfometria do relevo (Barbosa et al., 2011). Amostragem de solos com base nas diferenças do relevo é fundamental para estratificação da paisagem e é largamente utilizada em levantamentos pedológicos (Embrapa, 1995).

No Estado de São Paulo, as quadrículas na escala 1:100.000 de Brotas e Piracicaba possuem mapa pedológico (Almeida et al., 1981; Oliveira & Prado, 1989) e, assim, podem ser utilizadas como referência para o treinamento de dados para mapeamento digital de classes de solos e posterior aplicação em área semelhante que não possua o mapa. A folha de escala 1:50.000 de Botucatu não tem mapa pedológico em grande extensão; todavia, possui muitas semelhanças geomorfológicas, geológicas e climáticas com as quadrículas de Brotas e Piracicaba (IPT, 1981a,b), sugerindo que dados dessas áreas de referência possam ser utilizados no mapeamento da folha Botucatu.

Os objetivos deste estudo foram produzir um mapa pedológico para a folha Botucatu (SF-22-Z-B-VI-3) por meio de técnicas de mineração de dados aplicadas a variáveis geomorfométricas e de geologia, com base em áreas de referência; e testar a confiabilidade desse mapa por meio de validação em campo, com diferentes sistemas de amostragem.

MATERIAL E MÉTODOS

As áreas de estudo se inserem em três folhas cartográficas na escala 1:50.000: Dois Córregos (SF-22-Z-B-III-3), São Pedro (SF-23-Y-A-IV-1) e Botucatu (SF-22-Z-B-VI-3), na região central do Estado de São Paulo. Essas folhas delimitam-se pelas seguintes coordenadas geográficas: 48° 30' - 48° 15' W e 22° 15' - 22° 30' S (Dois Córregos), 48°00' - 47° 45' W e 22° 30' - 22° 45' S (São Pedro) e 48° 30' - 48° 15' W e 23° 00' - 23° 15' S (Botucatu). As folhas São Pedro e Dois Córregos já possuem mapa pedológico elaborado por métodos tradicionais, pertencente respectivamente às quadrículas Piracicaba (Oliveira & Prado, 1989) e Brotas (Almeida et al., 1981), de escala 1:100.000. A folha Botucatu (SP) não possui mapa pedológico.

De acordo com a classificação climática de Köppen, o clima nas cidades de Dois Córregos (677 m de altitude) e Botucatu (840 m de altitude) é do tipo Cwa, subtropical quente sem estação seca, com médias anuais de temperatura de 21 °C e precipitação pluvial de 1.350 mm. Na folha São Pedro (550 m), o clima é Aw, tropical com inverno seco, com médias anuais de temperatura a 22 °C e precipitação pluvial de 1.300 mm (Cepagri, 2012). Apesar de pequena diferenciação climática entre as três áreas de trabalho, o clima dessas foi considerado homogêneo para fins desse mapeamento, o que permitiu excluir o clima regional como fator preditivo dos solos.

Segundo o mapa geológico em escala 1:250.000, folha Bauru, SF 22-Z-B (DAEE/UNESP, 1984), as quatro formações geológicas mais representativas na área a ser mapeada (folha Botucatu esc. 1:50.000) são: formação Marília, do grupo Bauru, composta por arenitos grossos, imaturos, maciços, com abundantes nódulos calcíferos; formação Serra Geral, do grupo São Bento, constituída por basaltos toleíticos de coloração cinza a negro, intercalados por arenitos intertrapeanos; formação Botucatu, do grupo São Bento, constituída principalmente de arenitos eólicos; e formação Piramboia, do grupo São Bento, composta por arenitos finos a médios com matriz siltico-argilosa, estratificação cruzada de médio a grande porte.

O mapa geomorfológico do Estado de São Paulo na escala 1:1.000.000 (IPT, 1981a) mostra que o relevo na folha Botucatu constitui parte de três províncias geomorfológicas desse Estado: Depressão Periférica, no leste; Cuestas Basálticas; e Planalto Ocidental, no oeste, que são compartimentos com agrupamento de formas de relevo relativamente homogêneas, em relação aos demais compartimentos. A folha Dois Córregos também abrange parte dessas mesmas três províncias geomorfológicas, enquanto, na folha São Pedro, são encontradas formas de relevo representativas, principalmente da Depressão Periférica, com pequena porção nos domínios de relevo das Cuestas Basálticas e do Planalto Ocidental Paulista.

A análise dos mapas geológicos na escala 1:250.000, folhas Bauru (SF 22-Z-B) (DAEE/UNESP, 1984) e Campinas (SF 23-Y-A) (DAEE/UNESP, 1982), e do mapa geomorfológico do Estado de São Paulo na escala 1:1.000.000 (IPT, 1981a) permitiu verificar que a folha Botucatu (1:50.000), área foco deste mapeamento, e as folhas Dois Córregos (1:50.000) e São Pedro (1:50.000), áreas para treinamento dos algoritmos, são semelhantes quanto a geologia e relevo. Isso evidenciou a possibilidade de treinarem-se os algoritmos nas folhas Dois Córregos e São Pedro, para mapeamento dos solos da folha Botucatu.

A obtenção da base de dados foi uniforme para as três folhas cartográficas do estudo. As cartas topográficas 1:50.000 do Instituto Brasileiro de Geografia e Estatística (IBGE) e disponíveis em formato raster (<http://www.ibge.gov.br/home/geociencias/cartografia/default.shtm>) foram vetorizadas no programa ArcGIS, obtendo-se os planos de informação (PIs): malha viária, composto de ferrovias, caminhos, estradas de terra, estradas pavimentadas e rodovias; hidrografia, composto de rios e lagos/represas; hipsometria, que compreende curvas de nível e pontos cotados; e manchas urbanas, digitalizadas em polígonos.

O mapa temático de geologia (polígonos e linhas) na escala 1:500.000 (IPT, 1981b) e as cartas pedológicas das quadrículas de Brotas (Almeida et al., 1981) e Piracicaba (Oliveira & Prado, 1989), em escala

1:100.000, foram escaneados, georreferenciados e vetorizados no programa ArcGIS. Nessas cartas pedológicas, as legendas das unidades de mapeamento foram padronizadas até o 4º nível categórico do Sistema Brasileiro de Classificação de Solos mais agrupamento textural (Embrapa, 2006).

A partir dos dados de hipsometria das áreas de estudo, foram elaborados os modelos digitais de elevação (MDE) com resolução de 30 m. Com base nos atributos usados por Crivellini et al. (2009), selecionaram-se variáveis morfométricas para a composição do banco de dados: curvatura em perfil, curvatura em planta, distância diagonal e declividade. Os protocolos de treinamento foram completados com a seleção das variáveis: altitude, direção de fluxo e índice topográfico combinado (Moore et al., 1993; Irvin et al., 1997; Carvalho Júnior et al., 2003; Borges et al., 2005; Chagas, 2006; Valladares & Hot, 2006).

Em razão do tamanho muito grande do banco de dados conjunto das folhas Dois Córregos e São Pedro (cerca de 1,4 milhão de registros), impossibilitar a análise pelo programa de mineração de dados, foi realizada uma subamostragem aleatória de 50 % dos dados de cada uma dessas folhas, que foram reunidos gerando uma única base de dados contendo informações das duas áreas.

O treinamento dos dados foi feito com a utilização do algoritmo de árvore de decisão Part, presente no aplicativo de mineração de dados *Waikato Environment for Knowledge Analysis* (Weka) (Witten & Frank, 2005), que gera regras de classificação por meio de aprendizado de máquina. Foram usados 90 % dos registros da matriz de dados, escolhidos aleatoriamente, para treinamento pelo algoritmo e 10 %, para validar o modelo gerado. O treinamento em 90 % dos dados visou maior aprendizado pelo programa, sem prejuízo da validação do modelo, já que o banco de dados é muito extenso.

Para a geração do mapa pedológico digital, as regras foram divididas por unidade de mapeamento e posteriormente adaptadas ao formato requerido pelo programa de computador Ilwis (versão 3.7) por meio da lógica Boleana, criando-se um mapa para cada unidade de mapeamento. Por fim, fez-se a sobreposição dos mapas de unidades de mapeamento, o que possibilitou a obtenção dos mapas pedológicos digitais da folha Botucatu.

Testaram-se duas técnicas de validação em campo dos mapas: amostragem por transectos e amostragem estratificada. Na folha São Pedro, também utilizada para treinamento dos dados, foi selecionada a unidade de mapeamento PVAd arenosa/media (Argissolo Vermelho-Amarelo distrófico textura arenosa/média) do mapa pedológico digital, produzido com técnicas de árvores de decisão por Crivellini et al. (2009). Essa escolha foi em razão da grande representatividade da unidade - 52 % da área total da folha no mapa digital e 45 % da área da folha no mapa tradicional (Oliveira & Prado, 1989), com legenda simplificada - e do seu

alto grau de fragmentação no mapa digital (Crivelenti et al., 2009). Nessa unidade de mapeamento, selecionaram-se dois polígonos e, em cada um deles, foram feitos dois transectos de 5 km de extensão, com 25 pontos de amostragem distanciados entre si de 200 m, totalizando 100 pontos de observação. Nesses pontos, por meio de sondagens com trado até 2 m de profundidade, os solos foram caracterizados morfológicamente, de acordo com Santos et al. (2005), e coletadas amostras nos horizontes genéticos superficiais e subsuperficiais em alguns pontos do transecto, para determinação da granulometria, segundo Camargo et al. (1986), e caracterização química- pH em água e em CaCl_2 , matéria orgânica, P disponível, e Al, H+Al, K, Ca e Mg trocáveis, conforme Raij et al. (2001), calculando-se soma de bases (SB), capacidade de troca catiônica a pH 7 (CTC) e saturação por bases (V%). Os solos foram classificados de acordo com o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006), em nível de Subgrupo (4º nível) mais grupamento textural.

A validação do mapa digital da folha Botucatu foi feita com pontos amostrais distribuídos pela técnica do hiper-cubo latino, que permite aleatorização dentro dos diferentes padrões de associação solo-paisagem, ou seja, é uma amostragem aleatório-estratificada. Foi delimitado um *buffer* (bordadura) de 50 m ao longo da malha viária para facilitar o processo de amostragem e os dados numéricos das variáveis morfométricas e de geologia do banco de dados da folha Botucatu, na área delimitada por esse *buffer*, foram padronizados de acordo com o programa de computador, que realiza os sorteios pelo hiper-cubo latino, o cLHS (Wyss & Jorgensen, 1998). Esses dados foram analisados pelo cLHS, com 20.000 iterações, que selecionou 100 pontos amostrais de solo. O arquivo de texto contendo os pontos amostrais foi convertido para formato Excel e utilizado pelo ArcGIS (versão 10) e pela unidade de GPS. A caracterização morfológica dos solos nesses pontos foi feita de acordo com Santos et al. (2005) em sondagens com trado até 2 m e em minitrincheiras, coletando-se também outras informações da paisagem, como posição na vertente e declividade. Amostras dos horizontes superficial e subsuperficial de cerca de 1/3 do total de pontos foram analisadas para granulometria e química, bem como os solos classificados em nível de subgrupo (4º nível) mais grupamento textural, de acordo com o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006).

A classificação dos solos foi inserida em Sistema de Informação Geográfica (Ilwis 3.7) e esses pontos foram comparados com os mapas digitais, fazendo-se uso de matrizes de confusão. A exatidão foi avaliada por meio da “exatidão global”, proporção de observações corretamente classificadas em relação ao número total de observações; da “exatidão do produtor”, proporção de uma unidade de mapeamento classificada corretamente; da “exatidão do usuário”, probabilidade de um ponto no mapa representar a verdade de campo; do índice kappa, proporção da concordância observada

que se aproxima da concordância perfeita, retirando-se os efeitos do acaso; e de índices kappa condicionais, índices kappa para cada unidade de mapeamento. Os índices kappa foram analisados, segundo critério de Landis & Koch (1977).

RESULTADOS E DISCUSSÃO

A validação no campo do mapa da folha São Pedro por meio de transectos identificou sete classes de solos (Quadro 1), sendo duas delas de Argissolos Vermelho-Amarelos distróficos típicos, que diferem apenas pelo grupamento textural, arenosa/média e média/argilosa, e outras duas de Argissolos Vermelho-Amarelos distróficos textura arenosa/média, que diferem pela espessura do horizonte superficial arenoso (arênicos e típicos); essa última distinção só foi analisada na comparação entre as observações de campo e o mapa tradicional, pois a legenda simplificada do mapa digital não incluiu o caráter arênico.

Das 100 observações de solo em campo, 86 pertencem à subordem dos Argissolos Vermelho-Amarelos, mas, em nível de grande grupo mais grupamento textural, a exatidão global do mapa digital foi 83 % (Quadro 1), compatível com levantamento em nível de reconhecimento de alta intensidade (Embrapa, 1995). Como unidade de mapeamento, Argissolo Vermelho-Amarelo distrófico textura arenosa/média é a mais representativa da folha São Pedro (52 % da área), sugerindo boa confiabilidade para esse mapa digital. A proporção de 14 % das observações de campo não previstas no mapa digital, correspondente às classes Cambissolos, Gleissolos, Neossolos Flúvicos e Plintossolos, está dentro do limite de 15 % admitido como inclusão de solos; sua não predição deveu-se a essas classes de solo não constarem na legenda dos mapas das áreas de treinamento.

Os altos índices kappa condicional, calculados para as unidades de Argissolo (0,71 e 0,78) (Quadro 1), apresentaram nível de exatidão muito bom, o que se esperava, já que os transectos foram locados em unidades de mapeamento de Argissolos. Como os transectos tiveram traçado aproximadamente retilíneo e o mapa digital tem delineamentos bastante fragmentados, na prática, os transectos cruzam delineamentos de PVA aren/med e PVA med/arg. Entretanto, o índice kappa obtido, considerando todas as classes de solo identificadas no campo ($\text{kappa} = 0,28$), foi apenas razoável, já que muitas delas não aparecem no mapa digital.

Por causa da legenda mais detalhada do mapa tradicional, em relação ao mapa digital, a verificação da exatidão do mapa tradicional deu-se em nível de subgrupo mais grupamento textural (Quadro 2). Além do grupamento textural dos Argissolos, pode ser verificado no campo o critério espessura de horizonte superficial arenoso e com isso enquadrar os Argissolos

Quadro 1. Número de observações com solo classificado no campo (coluna) e no mapa digital (linha) na folha São Pedro

Unidade de mapeamento ⁽¹⁾	PVAd are/med	PVAd med/arg	CXbd	GX_GM	RYbd	FXd	Total	Exatidão do usuário	Kappa condicional
								%	
PVAd aren/med	79	1	4	5	1	1	91	86,8	0,71
PVAd med/arg	2	4	0	3	0	0	9	44,4	0,78
CXbd	0	0	0	0	0	0	0	-	0
GX_GM	0	0	0	0	0	0	0	-	0
RYbd	0	0	0	0	0	0	0	-	0
FXd	0	0	0	0	0	0	0	-	0
Total	81	5	4	8	1	1	100		
Exatidão do produtor (%)	97,5	80	0	0	0	0			

⁽¹⁾PVAd are/med: Argissolo Vermelho-Amarelo distrófico, textura arenosa/média; PVAd med/arg: Argissolo Vermelho-Amarelo distrófico, textura média/argilosa; CXbd: Cambissolo Háptico Tb distrófico, textura média; GX_GM: Gleissolos Hápticos e Melânicos indiscriminados; RYbd: Neossolo Flúvico Tb distrófico, textura indiscriminada; e FXd: Plintossolo Háptico distrófico, textura média.

Quadro 2. Número de observações com solo classificado no campo (coluna) e no mapa tradicional (linha) na folha São Pedro

Unidade de mapeamento ⁽¹⁾	PVAd are/med	PVAd med/arg	PVAd aren aren/med	CXbd	GX_GM	RYbd	FXd	Total	Exatidão do usuário	Kappa condicional
									%	
PVAd aren/med	42	1	5	4	5	1	1	59	71,2	0,36
PVAd med/arg	10	4	0	0	2	0	0	16	25	0,76
PVAd aren aren/med	3	0	19	0	0	0	0	22	86,4	0,73
CXbd	0	0	0	0	0	0	0	0	0	0
GX_GM	2	0	0	0	1	0	0	3	33,3	0
RYd	0	0	0	0	0	0	0	0	-	0
FXd	0	0	0	0	0	0	0	0	-	0
Total	57	5	24	4	8	1	1	100		
Exatidão do produtor (%)	73,7	80	79,2	0	12,5	0	0			
Exatidão global = 66 %; Kappa = 0,43										

⁽¹⁾ PVAd are/med: Argissolo Vermelho-Amarelo distrófico típico, textura arenosa/média; PVAd med/arg: Argissolo Vermelho-Amarelo distrófico típico, textura média/argilosa; PVAd aren aren/med: Argissolo Vermelho-Amarelo distrófico arênico, textura arenosa/média; CXbd: Cambissolo Háptico Tb distrófico típico, textura média; GX_GM: Gleissolos Hápticos e Melânicos indiscriminados; RYbd: Neossolo Flúvico Tb distrófico típico, textura indiscriminada; e FXd: Plintossolo Háptico distrófico típico, textura média.

em típico ou arênico (textura arenosa da superfície do solo até um mínimo de 50 cm e máximo de 100 cm de profundidade), feição comum na área do estudo, contemplada na legenda do mapa tradicional.

Os valores da exatidão global e do índice kappa do mapa tradicional apresentaram-se relativamente baixos (Quadro 2), pois esses levam em consideração toda a matriz de confusão, incluindo as classes pouco representativas, consideradas inclusões. Ainda assim, esses valores foram bem superiores aos do mapa digital, pois o mapa tradicional previu ocorrência de quatro unidades de mapeamento e por isso obteve maior exatidão global do mapeamento. Os valores de exatidão global ainda baixos devem-se ao fato de os transectos

de amostragem terem sido locados na unidade do mapa digital, cujo delineamento difere daquele do mapa tradicional. A mesma explicação aplica-se aos índices kappa condicional e de exatidão mais elevados da unidade Argissolos Vermelho-Amarelos distróficos típicos média/argilosa, unidade que apresentou número pequeno de acertos; porém, no balanço entre erros e acertos foram obtidos valores altos de exatidão.

Para a geração do mapa digital de classes de solos, folha Botucatu, produziram-se 206 regras pelo algoritmo PART, distribuídas conforme o quadro 3.

A unidade de mapeamento mais representada por regras foi a de Latossolo Vermelho-Amarelo distrófico

típico textura média com 21 regras (Quadro 3), com grande representatividade na matriz de dados (29 % da área total). O algoritmo conseguiu gerar regras para todas as unidades de mapeamento contidas na matriz de dados de treinamento e, mesmo unidades de mapeamentos pequenas, como a de Espodossolo Humilúvico órtico (seis regras), foram representadas por regras. Ainda assim, o mapa digital não apresentou todas as unidades de mapeamento das áreas de treinamento, já que algumas regras não são passíveis de aplicação na folha Botucatu, em razão de a área de treinamento ter variações no relevo e na geologia, não encontradas nessa folha.

Com a utilização do SIG Ilwis 3.7 e a partir das regras geradas pelo algoritmo PART, foi elaborado o mapa digital de solos da folha Botucatu (Figura 1). A extensão de ocorrência dos solos no mapa digital (Quadro 4) e os resultados da validação desse mapa (Quadro 5) também são apresentados. Foram previstas

Quadro 3. Número de regras produzidas pelo algoritmo PART e extensão relativa das unidades de mapeamento da área de treinamento

Unidade de mapeamento ⁽¹⁾	Área %	Regra
LVAdt_med	28,99	21
LVAdt_muitoarg	0,04	7
LVdf_arg_ou_muitoarg	1,92	3
LVAdt_arg	0,01	7
LVAet_muitoarg	0,03	5
LVdt_med	2,52	5
PVAdt_are/med	19,35	17
PVAdabrut_are/med	5,83	9
PVAaren_are/med	3,68	4
PVAarenabrup_are/med	10,32	12
NVdf_arg	2,67	6
NVdt_arg	0,85	4
NVef_arg	0,56	3
RLdt_arg	2,62	11
RLdt_med	4,95	11
RLet_arg	0,81	7
RYdt	0,07	9
RQo	12,85	16
MXo_arg	0,03	9
GX_GM_indisc	1,66	5
CXbd_are_ou_arg	0,19	4
EKo_aren	0,05	6

⁽¹⁾LVA: Latossolo Vermelho-Amarelo; LV: Latossolo Vermelho; PVA: Argissolo Vermelho-Amarelo; NV: Nitossolo Vermelho; RL: Neossolo Litólico; RY: Neossolo Flúvico; RQ: Neossolo Quartzarênico; MX: Chernossolo Háptico; GX_GM: Associação de Gleissolos Hápticos e Melânicos; CX: Cambissolo Háptico; EK: Espodossolo Humilúvico; d: distrófico; df: distrófico; o: órtico; abrup: abrupático; aren: arênico; t: típico; med: média; muitoarg: muito argilosa; arg: argilosa; are: arenosa; e indic: indiscriminada.

nove unidades de mapeamento pedológico. A distribuição dessas unidades, associada às variáveis predictoras de relevo e geologia, evidencia compatibilidade com a distribuição dos solos na paisagem estudada por outros autores em condições similares (Teramoto et al., 2001; Piroli, 2002; Sirtoli, 2008).

O índice de exatidão global obtido no mapeamento digital da folha Botucatu foi de 52 % e o índice kappa, de 0,41, este último indica qualidade Boa, segundo critério de Landis & Koch (1977). A exatidão global de 52 % foi baixa em relação a outros mapeamentos que utilizaram árvores de decisão para prever relações solo-paisagem, a exemplo da obtida por Zhou et al. (2004), 81,3 %, e por Hansen et al. (2009), 75,5 %. Da mesma forma, comparando-se a exatidão de 52 % com a precisão esperada em unidades de mapeamento (Embrapa, 1995), a precisão desse mapa é compatível com aquela esperada em nível de reconhecimento de baixa intensidade, que requer 50 a 70 % de confiabilidade, muito embora esse referencial seja para avaliar qualidade de unidades de mapeamento individualmente e não coletivamente. A rigor, a exatidão do mapa pedológico digital treinado em mapas tradicionais não deve ser maior que aquela dos mapas de treinamento. Apesar de não se ter avaliação ampla

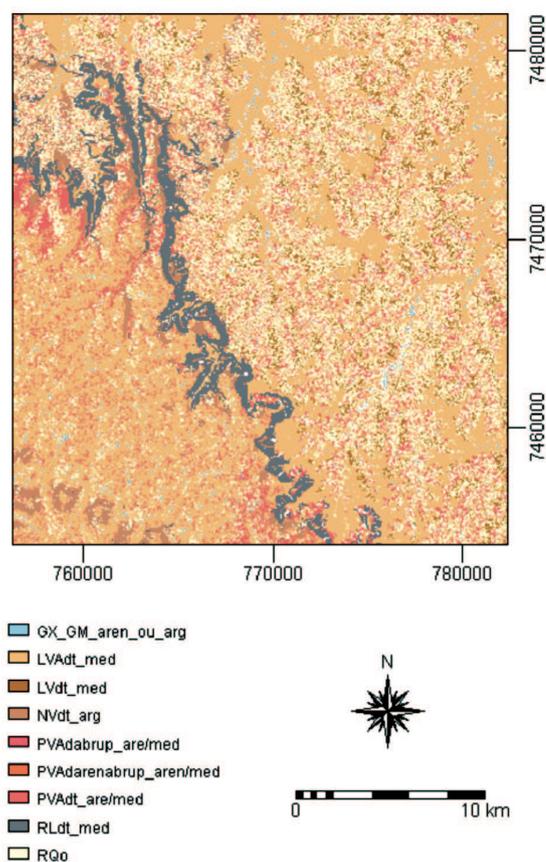


Figura 1. Mapa digital de solos da folha Botucatu, produzido a partir das variáveis geomorfométricas e litologia, por meio do algoritmo PART.

Quadro 4. Extensão de ocorrência das unidades de mapeamento do mapa pedológico digital da folha Botucatu (SF-22-Z-B-VI-3)

Unidade de mapeamento	Área relativa na folha
	%
Latossolo Vermelho-Amarelo distrófico típico textura média	49
Neossolo Quartzarênico órtico	23
Argissolo Vermelho-Amarelo distrófico típicotextura arenosa/média	11
Argissolo Vermelho-Amarelo distrófico arênico abrupto textura arenosa/média	1
Argissolo Vermelho-Amarelo distrófico abrupto textura arenosa/média	1
Latossolo Vermelhadistrófico típico textura média	4
Nitossolo Vermelho distrófico típico textura argilosa	4
Neossolo Litólico distrófico textura média	6
Gleissolo de textura indiscriminada	1

da qualidade dos mapas de treinamento usados (folhas São Pedro e Dois Córregos), a validação de campo da unidade PVAd arenosa/média da folha São Pedro evidenciou confiabilidade compatível com mapeamentos em nível de reconhecimento de baixa intensidade (exatidão global = 66 %). Assim, estendendo-se a confiabilidade dessa unidade de mapeamento (PVAd a/m), que ocupa mais da metade da área do mapa digital da folha São Pedro, para toda a folha, a exatidão global de 52 % encontrada no mapa pedológico digital da folha Botucatu é coerente com a confiabilidade do mapa de treinamento da de São Pedro, aceitável para mapas de reconhecimento de baixa intensidade (Embrapa, 1995). Perda adicional de acurácia do mapa digital (66 para 52 %) deve-se ao treinamento dos dados - e a elaboração das regras - ter sido executado a partir de mapas tradicionais e não em observações pontuais realizadas *in situ*, o que seria mais adequado por excluir eventuais erros inerentes à elaboração dos mapas de treinamento.

O alto índice kappa indica alto poder preditivo do método com relação a unidades de mapeamento mais comuns na área de estudo, como Latossolo Vermelho-Amarelo distrófico típico textura média. Em contrapartida, parte dos erros de classificação do mapa digital foi em razão de que as unidades de mapeamento menores foram confundidas com as maiores, ou seja, atributos das variáveis de treinamento (geomorfométricos e de geologia) das unidades de mapeamentos maiores, com maior amplitude de variação, são mais amplamente reproduzidos e, assim, confundidos com padrões de unidades menores. Exemplo disso é o Latossolo Vermelho distrófico típico textura média (unidade de menor tamanho) que foi muito predito como outras unidades de maior tamanho, como Latossolo Vermelho-Amarelo distrófico típico textura média e Argissolo Vermelho-Amarelo distrófico típico textura média. Parte dos erros, também se deve ao fato de que há classes de solos encontradas no campo que não estão presentes no mapa de treinamento, como as de Planossolos e Chernossolos.

Foi obtido índice kappa condicional de 0,63 para a classe Argissolo Vermelho-Amarelo distrófico típico

textura arenosa/média, classe de qualidade Muito Boa (Landis & Koch, 1977). Obtiveram-se ainda os índices kappa condicional de 0,53 para a classe Nitossolo Vermelho distroférrico típico; 0,56, para a classe Neossolo Quartzarênico órtico, ambos da classe de qualidade Boa; e 0,33, para a classe Latossolo Vermelho distrófico textura média, classe de qualidade Razoável. O restante das classes apresentou índice kappa inferior a 0,2, classe de qualidade Ruim. Como essas classes também foram pouco representadas na matriz de treinamento, evidenciou-se o baixo poder preditivo do método para unidades de mapeamento pequenas, pois se geram menos regras e os padrões de ocorrência desses solos não são representados.

Os altos índices kappa condicionais encontrados para as unidades de Argissolo Vermelho-Amarelo distrófico típico textura arenosa/média retratam o alto grau de concordância do mapa digital de solos com os dados reais. Essa unidade de mapeamento é a segunda mais extensa da matriz de dados e a segunda com maior número de regras geradas (17), portanto com padrões de associação solo-relevo bem representados. Um ponto foi classificado como Neossolo Litólico distrófico, o que concordou com sua classificação no mapa; por isso, a exatidão do usuário para essa unidade é 100 %.

Apesar de muitas unidades terem índice kappa condicional alto, o índice kappa obtido, considerando todas as classes de solo identificadas no campo (kappa=0,41), apesar de bom, foi menor que os condicionais, pois muitas unidades não aparecem no mapa digital. As unidades de mapeamento encontradas no campo e não preditas no mapa digital de solos são 17 % do total e podem ser consideradas inclusões de solos, pois esse valor está dentro do limite de 20 % de observações de solos distintos admitidas como inclusões em mapas em nível de reconhecimento (Embrapa, 1995). Porém, esse critério é usado por unidade de mapeamento e não para um mapa composto de várias unidades, reforçando a necessidade também de índices que avaliem os mapas como um todo, e não só índices por unidade de mapeamento. Sob esse aspecto, a validação aplicada na folha São Pedro, por unidade de

Quadro 5. Número de observações com solo classificado no campo (coluna) e no mapa digital (linha) da folha Botucatu (SF-22-Z-B-VI-3)

Unidade de mapeamento ⁽¹⁾	GX GM_are_ ou arg	NVd targ	PVAd abrup are/med	PVAdt aren are/med	PVA aren abrup are/med	PVAdt are/med	PVAdt_ med/arg	MXo arg	LVAd t med	LVdt arg	LVdt med	LVef arg	RLdt med	RQo med	RRdt	FFc	Total	Exat. usuário
GX_GM_are_ou_arg	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0,0
NVdt_arg	1	4	0	0	0	1	0	0	1	1	0	0	0	1	0	0	9	44,4
PVAd abrup are/med	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
PVAdt aren are/med	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
PVA aren abrup are/med	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PVAdt are/med	0	0	3	1	2	14	2	0	1	2	3	0	0	1	0	0	29	48,3
PVAdt med/arg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
MXo arg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
LVAdt med	1	3	0	0	0	2	0	1	19	4	4	1	0	4	0	1	40	47,5
LVdt arg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
LVdt_med	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	4	100,0
LVef_arg	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
RLdt_med	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	3	33,3
RQo	0	0	0	0	0	2	0	0	2	0	0	0	0	10	0	0	14	71,4
RRdt_med	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
FFc	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
Total	2	7	3	1	2	19	2	1	24	7	11	1	1	16	2	1	100	
Exatidão do produtor	0,0	57,1	0,0	0,0	0,0	73,7	0,0	0,0	79,2	0,0	36,4	0,0	100,0	62,5	0,0	0,0		

Exatidão global = 52 %; Índice Kappa = 0,41

⁽¹⁾LVAdt_med: Latossolo Vermelho-Amarelo distrófico típico textura média; RQo: Neossolo Quartzarênico órtico; PVAdt_are/med: Argissolo Vermelho-Amarelo distrófico típico textura arenosa/média; PVAdt_med/arg: Argissolo Vermelho-Amarelo distrófico típico textura média/argilosa; MXo: Chernossolo Háptico órtico típico textura argilosa; PVAdt aren are/med: Argissolo Vermelho-Amarelo distrófico arênico textura arenosa/média; PVAd aren abrup_are/med: Argissolo Vermelho-Amarelo distrófico arênico abrupto textura arenosa/média; PVAd abrup_are/med: Argissolo Vermelho-Amarelo distrófico abrupto textura arenosa/média; LVdt_arg: Latossolo Vermelho distrófico típico textura argilosa; LVdt_med: Latossolo Vermelho distrófico típico textura média; LVef_arg: Latossolo Vermelho eutroférico típico textura argilosa; NVdt_arg: Nitossolo Vermelho distrófico típico textura argilosa; RLdt_med: Neossolo Litólico distrófico textura média; GX_GM_are_ou_arg: Gleissolos de textura indiscriminada; RRdt med: Neossolo Regolítico distrófico típico; e FFc: Plintossolo Pétrico concrecionário léptico.

mapeamento, enquadra-se melhor nos critérios de avaliação de qualidade utilizados para mapeamentos tradicionais.

A unidade Latossolo Vermelho-Amarelo distrófico textura média foi contemplada no mapa digital com 40 pontos de amostragem; desses, apenas 19 tiveram classificação no campo coincidente com a legenda do mapa (“exatidão do usuário” 47,5 %). O índice kappa desse Latossolo foi de 0,3, considerado Razoável, pois apesar de essa unidade ter sido muito representada no treinamento dos dados e, assim, terem sido geradas muitas regras (21), menos da metade de sua ocorrência foi validada no campo como outra classe, o que pode ter acontecido pela existência de padrões de associação solo-paisagem na área de treinamento que não se reproduziram na área de teste (folha Botucatu), ou seja, apesar da grande ocorrência desse tipo de solo na folha Botucatu, os padrões de ocorrência são distintos.

A unidade Neossolo Quartzarênico órtico típico apresentou índice kappa condicional Bom e “exatidão do produtor” (62,5 %) e “do usuário” (71,4 %) elevados, ou seja, essa unidade de mapeamento foi muito representativa na área de treinamento e, assim, muito representada por regras. Porém, ao contrário do que ocorreu com a unidade de mapeamento LVAdt_med, os padrões solo-relevo no treinamento coincidiram com a área de teste, o que pode ser explicado pela forte associação dessa unidade de mapeamento à formação geológica Piramboia, ressaltando o alto poder preditivo da variável geologia para essa unidade.

A “exatidão do usuário” para a unidade Argissolo Vermelho-Amarelo distrófico típico textura arenosa/média manteve-se elevada (76 %), evidenciando que o poder preditivo do método em relação aos dados de campo é alto. A probabilidade de que essa unidade representada no mapa realmente corresponda à verdade terrestre é bastante alta, ao contrário da “exatidão do produtor” (48,3 %), que expõe que essa unidade de mapeamento teve baixo número de pontos classificados corretamente no mapa digital, ou seja, ela foi muito confundida, principalmente com unidades de Argissolos diferentes no nível de subgrupo. Pode-se notar também que em todas as unidades de mapeamento pertencentes à ordem dos Argissolos, 100 % dos pontos coincidiram com a unidade Argissolo Vermelho-Amarelo distrófico típico textura arenosa/média.

A unidade de mapeamento Argissolo Vermelho-Amarelo distrófico típico textura arenosa/média ocorre em grande extensão, assim como a unidade Latossolo Vermelho-Amarelo distrófico típico textura média. Entretanto, ao contrário desta última, apresentou elevada exatidão, em razão de a unidade de Argissolos estar mais diretamente ligada à ocorrência da formação Piramboia que a de Latossolos. Os padrões de relevo onde ocorrem os Latossolos também são mais propensos à ocorrência de outras classes de solos, como os Neossolos Quartzarênicos.

A unidade de mapeamento Nitossolo Vermelho distroférico típico apresentou “exatidão do usuário” de 44,4 % e “exatidão do produtor” de 57,1 %. Isso ocorreu, pois a maioria dos pontos encontrados no campo com essa classificação foi identificada corretamente no mapa; porém, poucos pontos com essa classificação no mapa coincidiram com os dados de campo. Esses resultados indicam que as regras criadas incluem ambientes em que não ocorrem Nitossolos, ou seja, são muito abrangentes, pois existe baixa representatividade das unidades de mapeamento de solos derivados de basalto, como Nitossolo Vermelho distroférico típico e Latossolo Vermelho (distro/eutro) férico típico nas áreas de treinamento, o que resultou em baixo poder preditivo para essas unidades de mapeamento, produzindo baixa acurácia geral do mapa digital de solos.

Todos os pontos preditos pelo mapa digital como Latossolo Vermelho distrófico típico textura média coincidiram com os dados de campo (“exatidão do produtor” 100 %), enquanto poucos pontos de campo classificados como essa unidade foram identificados corretamente pelo mapa digital, evidenciando que para essa unidade de mapeamento o conjunto de dados de treinamento está bem representado na área de teste. Todavia, a variabilidade de padrões solo-relevo nas áreas de treinamento, para essa unidade de mapeamento, é pequena ao comparar-se com a área de teste, o que resultou em elevada “exatidão do usuário”, mas baixa “exatidão do produtor”.

As duas técnicas testadas de validação em campo dos mapas, por transectos e aleatório-estratificada, apresentaram uma diferença fundamental na aplicação do método; a primeira foi aplicada a uma unidade de mapeamento e, a segunda, a toda a área mapeada, ambas utilizando o mesmo número de pontos amostrais (100). A aplicação da técnica dos transectos para validação do mapeamento como um todo tornaria bastante difícil a cobertura de toda a extensão mapeada ou apresentaria esses com distribuição bastante descontínua, tornando essa espacialização mais assemelhada àquela por pontos aleatórios ou a amostragens puramente estratificadas. Além da dificuldade para sua locação nas diferentes unidades, a distribuição de vários transectos por toda a área tornaria a verificação operacionalmente mais trabalhosa, por dificultar o acesso e deslocamento em toda a área, aumentando também a duração do trabalho. Complementarmente, a técnica de amostragem pelo hipercubo latino facilita a distribuição da amostragem nos diversos padrões de associações solo-paisagem da área estudada, o que lhe confere elevada representatividade. Entretanto, a amostragem aleatório-estratificada só pode ser realizada quando se dispõe de uma base de dados compatível (topográficos, geológicos, climáticos etc.) para análise pelo *software* (cLHC), o que nem sempre está disponível nos trabalhos de levantamento. O entendimento das relações solo-paisagem por meio de pedossequências (topográficas,

climáticas, litológicas) também é facilitado pelo método dos transectos e não possível de estruturar pelo hipercubo latino, que usa critérios probabilísticos para estratificar e aleatorizar os pontos de amostragem.

CONCLUSÕES

1. A unidade de mapeamento no treinamento, quanto mais representativa for, maior será o número de regras geradas e melhor será a reprodução dos padrões solo-paisagem na área a ser mapeada. Por sua vez unidades de mapeamento de solos menores nas áreas de treinamento, com baixa representatividade, tendem a produzir menor número de regras, em alguns casos insuficientes para representar esses solos nas novas áreas a serem mapeadas.

2. A validação por transectos, utilizada em uma unidade de mapeamento do mapa digital da folha São Pedro, a unidade PVAd arenosa/média, indicou boa confiabilidade desse mapa, compatível com levantamentos de reconhecimento de alta intensidade, enquanto a confiabilidade do mapa tradicional dessa folha, após simplificação da legenda, mapa utilizado para treinamento, é compatível com levantamentos de reconhecimento de baixa intensidade.

3. A redução de exatidão do mapa pedológico digital da folha Botucatu (52 %), em relação à exatidão observada em mais da metade da folha São Pedro (66 %), usada para treinamento do algoritmo, foi relativamente baixa, provavelmente em razão de o treinamento ter sido feito em mapas e não em observações pontuais.

4. O método de amostragem aleatório-estratificada pelo hipercubo latino se adequou a trabalhos de mapeamento com base de dados extensa, para avaliar mapeamentos de forma ampla, tornando os trabalhos de campo nessas áreas operacionalmente mais eficientes. Já a amostragem em transectos foi compatível para avaliar a pureza de unidades de mapeamento individualmente, não necessitando de base de dados detalhada e permitindo estudos de associações solo-paisagem em pedossequências.

AGRADECIMENTOS

À Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de mestrado do primeiro autor. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio concedido por meio do projeto “Solos Frágeis” (CNPq no. 562698/2010-8). A todos que contribuíram no suporte técnico e na execução dos trabalhos de campo.

LITERATURA CITADA

- ALMEIDA, C.L.F.; OLIVEIRA, J.B. & PRADO, H. Levantamento pedológico semidetalhado do Estado de São Paulo: Quadrícula de Brotas (SF-22-Z-B-III). Campinas, Instituto Agrônomo de Campinas, 1981. Mapa, escala 1:100.000.
- BARBOSA, A.M.; PINHEIRO, H.S.K.; ANJOS, L.H.C.; CARVALHO JUNIOR, W. & CHAGAS, C.S. Determinação de pontos amostrais através de atributos do terreno para mapeamento digital dos solos da bacia hidrográfica do Rio Guapi-Macacu, RJ. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15., Curitiba, 2011, Anais... São José dos Campos, INPE, 2011. p.3672-3679.
- BERTOLANI, F. Variabilidade de atributos do solo em unidades de mapeamento de um levantamento pedológico semidetalhado. Campinas, Universidade Estadual de Campinas, 2003. 128p. (Tese de Doutorado)
- BORGES, M.E.S.; ARCOVERDE, G.F.B.; MARTINS, E.S.; RAMOS, V.M.; GUIMARÃES, R.F. & CARVALHO JÚNIOR, O.B. Mapeamento pedológico em relevos cársticos a partir da análise morfométrica. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., Goiânia, 2005. Anais...São José dos Campos, INPE, 2005. p.1725-1732.
- CAMARGO, O.A.; MONIZ, A.C.; JORGE, J.A. & VALADARES, J.M.A.S. Métodos de análise química e física de solos do Instituto Agrônomo de Campinas. Campinas, Instituto Agrônomo de Campinas, 1986. 94p. (Boletim Técnico, 106)
- CARVALHO JÚNIOR, O.A.; CARVALHO, A.P.F., GUIMARÃES, R.F.; LOPES, R.A.S.; GUIMARÃES, P.H.; MARTINS, E.S. & PEDRENO, J.N. Classification of hyperspectral image using SCM methods for geobotanical analysis in the brazilian savanna region. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 4., Toulouse, 2003. Proceedings... Piscataway, IEEE, 2003. p.3754-3756.
- CEPAGRI. 2012. Clima dos Municípios Paulistas. Disponível em <http://www.cpa.unicamp.br/outras-informacoes/clima-dos-municipios-paulistas_020307.html>. Acesso em: 15 de maio de 2012.
- CHAGAS, C.S. Mapeamento digital de solos por correlação ambiental e redes neurais em uma bacia hidrográfica no domínio de mar de morros. Viçosa, MG, Universidade Federal de Viçosa, 2006. 223p. (Tese de Doutorado)
- CRIVELANTI, R.C.; COELHO, R.M.; ADAMI, S.F. & OLIVEIRA, S.R.M. Mineração de dados para inferência de relações solo-paisagem em mapeamentos digitais de solo. *Pesq. Agropec. Bras.*, 44:1707-1715, 2009.
- DAEE/UNESP. Mapa Geológico do Estado de São Paulo: Folha Bauru (SF 22-Z-B). Rio Claro, DAEE/UNESP, 1984. Mapa, escala 1:250.000.
- DAEE/UNESP. Mapa Geológico do Estado de São Paulo: Folha Campinas (SF 23-Y-A). Rio Claro, DAEE/UNESP, 1982. Mapa, escala 1:250.000.

- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Procedimentos normativos de levantamentos pedológicos. Brasília, 1995. 116p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. 2.ed. Rio de Janeiro, Embrapa Solos, 2006. 316p.
- HAN, J. & KAMBER, M. Data mining: Concepts and techniques. Nova York, Morgan Kaufmann, 2001. 708p.
- HANSEN, M.K.; BROWN, D.J.; DENNISON, P.E.; GRAVES, S.A. & BRICKLEMYER, R.S. Inductively mapping expert-derived soil-landscape units within Dambo Wetland catenae using multispectral and topographic data. *Geoderma*, 150:72-84, 2009.
- INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO- IPT. Mapa Geomorfológico do Estado de São Paulo. v.1 (Nota Explicativa) e v.2 (Mapa). São Paulo, 1981a. Mapa, escala 1:1.000.000. (Série Monografias, 5)
- INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO - IPT. Mapa Geológico do Estado de São Paulo. v.1 (Nota Explicativa) e v.2 (Mapa). São Paulo, 1981b. Mapa, escala 1:500.000 (Série Monografias, 6)
- IRVIN, B.J.; VENTURA, S.J. & SLATER, B.K. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Alley. *Geoderma*, 77:137-154, 1997.
- LAGACHERIE, P. & VOLTZ, M. Predicting soil properties over a region using sample information from a mapped reference area and digital elevation data: A conditional probability approach. *Geoderma*, 97:187-208, 2000.
- LANDIS, J.R. & KOCH, G.G. The measurement of observer agreement for categorical data. *Biometrics*, 33:159-174, 1977.
- McBRATNEY, A.B.; MENDONÇA SANTOS, M.L. & MINASNY, B. On digital soil mapping. *Geoderma*, 117:3-52, 2003.
- MINASNY, B. & McBRATNEY, A.B. Uncertainty analysis for pedotransfer functions. *Eur. J. Soil Sci.*, 53:417-430, 2002.
- MOORE, I.D.; GESSLER, P.E.; NIELSEN, G.A. & PETERSON, G.A. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.*, 57:443-452, 1993.
- OLIVEIRA, J.B. & PRADO, H. Carta pedológica semidetalhada do estado de São Paulo: Folha Piracicaba (SF 23-Y-A-IV). Campinas, Instituto Agronômico de Campinas, 1989. Mapa, escala 1:100.000.
- PIROLI, E.L. Geoprocessamento na determinação da capacidade e avaliação do uso da terra do município de Botucatu (SP). Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho", 2002. 122p. (Tese de Doutorado)
- RAIJ, B.van.; ANDRADE, J.C.; CANTARELLA, H. & QUAGGIO, E.J.A. Análise química para avaliação da fertilidade de solos tropicais. Campinas, Instituto Agrônomo de Campinas, 2001. 285p.
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H.C. Manual de descrição e coleta de solo no campo. 5.ed. Viçosa, MG, Sociedade Brasileira de Ciência do Solo/Rio de Janeiro, Embrapa, 2005. 100p.
- SILVA, E.F. Comparação de mapas de solos produzidos em escalas e épocas distintas. Piracicaba, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 2000. 177p. (Tese de Doutorado)
- SIRTOLI, A.E. Mapeamento de solos com auxílio de atributos do terreno, índices espectrais e geologia integrados por redes neurais artificiais. Curitiba, Universidade Federal do Paraná, 2008. 102p. (Tese de Doutorado)
- TERAMOTO, E.R.; LEPSCH, I.F. & VIDAL-TORRADO, P. Relações solo, superfície geomórfica e substrato geológico na microbacia do ribeirão Marins (Piracicaba-SP). *Sci. Agríc.*, 58:361-371, 2001.
- VALLADARES, G.S. & HOTT, M.C. GIS and geomorphometry to pedological mapping. In: GLOBAL WORKSHOP ON DIGITAL SOIL MAPPING, 2., Rio de Janeiro, 2006. Anais... Rio de Janeiro, Embrapa Solos. CD-ROM
- WILDING, L.P. Spatial variability: Its documentation, accommodation and implication to soil surveys. In: NIELSEN, D.R. & BOUMA, J., eds. Soil spatial variability. WORKSHOP ISSS AND SSSA, LAS VEGAS (NV), Wageningen, 1984. Proceedings... Wageningen, PUDOC, 1985.
- WITTEN, I.H. & FRANK, E. Data mining: practical machine learning tools and techniques. 2.ed. San Francisco, Morgan Kaufmann, 2005. 524p.
- WYSS, G.D. & JORGENSEN, K.H. A user's guide to LHS: Sandia's Latin Hypercube Sampling Software. Albuquerque, Sandia National Laboratories, 1998. 138p. (SAND98-0210)
- ZHOU, B.; ZHANG, X. & WANG, R. Automated soil resources mapping based on decision tree and Bayesian predictive modeling. *J. Zhejiang Univ. Sci.*, 5:782-795, 2004.