

XXXII Congresso Brasileiro de Ciência do Solo

“Mineração de dados geomorfométricos em mapeamento digital de solos”

RAFAEL CASTRO CRIVELANTI⁽¹⁾, RICARDO MARQUES COELHO⁽²⁾, SAMUEL FERNANDO ADAMI⁽³⁾, STANLEY ROBSON DE MEDEIROS OLIVEIRA⁽⁴⁾ & MARCIO KOITI CHIBA⁽⁵⁾

RESUMO - O levantamento pedológico é a caracterização e classificação das unidades de solos, estabelecendo limites geográficos. Essa atividade é extremamente baseada em atividades de campo, o que torna os levantamentos onerosos e demorados. O mapeamento digital de solos é uma alternativa que pode torná-los mais acessíveis. O objetivo desta pesquisa foi desenvolver uma técnica de mapeamento digital de solos com apoio de parâmetros geomorfométricos, geologia e de base de dados de solos pré-existentes. A partir do Modelo Digital de Elevação, obteve-se os parâmetros geomorfométricos declividade, curvaturas em planta e perfil, área de contribuição da bacia e distância diagonal da drenagem da folha topográfica de São Pedro, escala 1:50.000. Esses parâmetros, juntamente com a geologia, foram associados, por georreferenciamento, aos mapas de solos, o que permitiu a construção de uma matriz relacional solos-geomorfometria. Essa matriz relacional foi utilizada para geração de um modelo de predição de solos, através da análise por árvores de decisão. Geologia foi o elemento com maior poder preditivo das unidades de mapeamento, seguido da variável declividade e distância diagonal da drenagem e, com acurácia de predição de 51%.

Palavras-Chave: (pedometria, levantamento, geologia; MDE; topografia)

Introdução

Segundo Santana [1], o levantamento pedológico é considerado como a caracterização e classificação das unidades de solos, estabelecendo limites geográficos, representados em um mapa, que apresenta sua descrição e sua interpretação com a finalidade proposta.

Esses inventários têm como principal objetivo subdividir áreas heterogêneas em parcelas homogêneas, de modo que apresentem a menor variabilidade possível, em função dos critérios de classificação dos solos e dos atributos utilizados para sua distinção [2]. No entanto, o levantamento tradicional é uma atividade bastante onerosa [3].

O mapeamento digital de solos introduz algumas vantagens em relação ao método tradicional, por ser uma alternativa rápida e econômica [4]. Pode ser definido como a criação de sistemas de informação espacial, utilizando modelos numéricos para a inferência das variações espaços-temporais dos tipos de solos e de suas propriedades, a partir de observações de variáveis ambientais correlacionadas [5].

A mineração de dados é uma técnica que pode ser utilizada para relacionar essas variáveis, através do processo exploratório em grandes bancos de dados à procura de padrões consistentes, permitindo o estabelecimento de novos subconjuntos [6].

Esse conhecimento pode ser analisado por ferramentas de árvores de decisão, criando-se a partir de registros conhecidos um conjunto de treinamento. Dessa árvore, pode-se classificar a amostra desconhecida sem necessariamente testar todos os valores dos seus atributos [7].

O objetivo deste trabalho foi desenvolver uma metodologia para o mapeamento digital de solos da folha São Pedro, estado de São Paulo, escala 1:50.000, com apoio de parâmetros geomorfométricos e de base de dados oriundos de levantamentos de solos existentes, através de técnicas de mineração de dados.

Material e Métodos

A. Área de estudo

A área deste estudo localiza-se na região central do estado de São Paulo, caracterizada pelo clima tropical com estação seca mais ampla no inverno. A precipitação média anual fica compreendida entre 1000 e 1500 mm. O relevo local tem elevada variabilidade e representa três províncias geomorfológicas [8]: a) Planalto Ocidental, com geologia da formação Itaqueri; b) Depressão Periférica, com ocorrências geológicas das formações Pirambóia, Corumbataí e de sedimentos aluviais; e c) Cuestas Basálticas, presente nas regiões de contato entre as províncias Planalto Ocidental e Depressão Periférica, onde ocorrem as formações geológicas Botucatu e Serra Geral (Almeida [9]).

⁽¹⁾ Primeiro Autor é Mestre em Agricultura Tropical e Subtropical, Centro de Solos e Recursos Ambientais (CSRA), Instituto Agronômico - IAC. Av. Barão de Itapura n.º 1481, Botafogo, CEP 13012-970, Campinas, SP. Email: grilasso@hotmail.com. ⁽²⁾ Segundo Autor é Pesquisador do CSRA/IAC, Av. Barão de Itapura n.º 1481, Botafogo, CEP 13012-970, Campinas, SP. Email: rmcoelho@iac.sp.gov.br. ⁽³⁾ Terceiro Autor é Pesquisador do CSRA/IAC, Av. Barão de Itapura n.º 1481, Botafogo, CEP 13012-970, Campinas, SP. Email: samuel@iac.sp.gov.br. ⁽⁴⁾ Quarto Autor é Pesquisador da Embrapa Informática Agropecuária, Av. André Tosello n.º 209, CEP 13.083-886, Barão Geraldo, Campinas, SP. Email: stanley@cnptia.embrapa.br. ⁽⁵⁾ Quinto Autor é Pesquisador do CSRA/IAC, Av. Barão de Itapura n.º 1481, Botafogo, CEP 13012-970, Campinas, SP. Email: mkchiba@iac.sp.gov.br.

B. Elaboração do banco de dados

A carta topográfica 1:50.000 em formato *raster* foi digitalizada e vetorizada. A carta geológica 1:1.000.000 [8] foi escaneada, georreferenciada e vetorizada. Da carta pedológica da quadrícula Brotas, escala 1:100.000 [9], previamente digitalizada, extraiu-se e gerreferenciou-se apenas a folha São Pedro. A legenda de solos do mapa pedológico original foi simplificada para a folha, unificando-se as unidades de mapeamento pelo 3º nível categórico do Sistema Brasileiro de Classificação de Solos [10] e por agrupamento de textura. O modelo digital de elevação (MDE) com 30 m de resolução foi gerado em ArcGIS [11] e os parâmetros geomorfológicos declividade, curvaturas em planta e perfil, distância diagonal da drenagem [12] e área de contribuição foram obtidos no software Ilwis Academic [13] e Idrisi Andes [14]. Após classificação dos mapas dessas variáveis de relevo, esses mapas foram cruzados com os mapas de geologia e solos da área, o que permitiu obtenção de uma matriz de dados com 794.273 linhas, em que cada linha representou um pixel (30x30m) do mapa 1:50.000, contendo informações discretas de cada parâmetro.

No software Weka 3.5.6 [15], após pré-processamento da matriz de dados, foram usados 90 % dos registros (linhas) da matriz de dados, escolhidos aleatoriamente, para treinamento por árvores de classificação e 10 % dos registros foram usados para validar o modelo gerado, ou seja, para verificar a acurácia ou poder preditivo da metodologia. Essa verificação foi baseada no coeficiente de incerteza da informação, com o qual obteve-se as acurácias geral do modelo e individuais de cada unidade de mapeamento na folha. O treinamento por árvores de classificação foi realizado em três diferentes balanceamentos de classes, recurso usado para não favorecer, na geração do modelo de aprendizado, as unidades de mapeamento com maior área de ocorrência. Os balanceamentos utilizados foram 0, 0,5 e 1, que representam, respectivamente, a distribuição original dos dados, a distribuição com subamostragem das classes (unidades de mapeamento) com maior ocorrência e a distribuição com igual proporção de ocorrência para todas as classes, em que as duas últimas são baseadas na técnica de amostragem com reposição.

Adicionalmente, visando obter um ordenamento dos parâmetros morfométricos para o modelo, foram realizados testes de ganho de informação por entropia e por comparação entre as frequências esperadas e observadas (teste χ^2).

Resultados

O resultado da análise utilizando árvore de decisão para os diferentes balanceamentos de classes, com todas as unidades de mapeamento (UNIMAPS) e após a retirada das unidades com probabilidade nula de ocorrência é apresentado na Tabela 1. A acurácia geral do modelo diminuiu com o aumento do balanceamento das UNIMAPS para os dois casos, pois na medida em

que esse é realizado, unidades de mapeamento que antes não constavam no modelo passaram a ser representadas. Isto resultou numa acurácia geral menor devido ao aumento das classificações incorretas imposto pela sub-amostragem das unidades de maior área de ocorrência. Este procedimento levou a um aumento na acurácia de 15,1% para 33,4%, com balanceamento igual a 1. Porém, o melhor desempenho foi verificado com o balanceamento de classes igual a 0, resultando num aumento na acurácia de 54,2% para 54,8%.

O balanceamento das classes apresentou resultado efetivo aumentando a acurácia em UNIMAPS que foram super-amostradas devido ao aumento do balanceamento de classes. Já a acurácia das unidades de mapeamento com grande representatividade diminuiu devido à deflação que o balanceamento de classes condiciona. Unidades de mapeamento que foram desconsideradas do modelo com balanceamento de classes igual a zero, como o Neossolo Lítico (RL) eutrófico textura média, passaram a ser consideradas e com uma acurácia de aproximadamente 30%, com o balanceamento de classes igual a 0,5. Outras com maior representatividade, como os Neossolos Quartzarênicos, tiveram sua acurácia diminuída com o balanceamento de classes passando de zero para 1, de 50% para 36% (Tabela 2).

Após a eliminação dos dados com probabilidade nula de ocorrência para o balanceamento de classes igual a 0, pode-se observar que a acurácia individual das unidades de mapeamento não apresentou mudança significativa (Tabelas 2 e 3). Há destaque para o RL eutrófico ou distrófico com textura média ou argilosa, que foi desconsiderada no balanceamento de classes igual a 0, porém passou a fazer parte do modelo nos balanceamentos das classes iguais a 0,5 e 1, com acurácia de aproximadamente 14,5%. Essa alteração na acurácia geral e individual das unidades de mapeamento no modelo ocorreu devido à baixa representatividade dessas nove unidades retiradas, por representarem menos de 1,5% da folha de estudo.

O ordenamento de fatores mais influentes na predição das unidades de mapeamento, em ordem decrescente, de acordo o teste χ^2 , foi: *geologia* > *declividade* > *distância diagonal da drenagem* > *área de contribuição da bacia* > *curvatura em perfil* > *curvatura em planta*. A importância das variáveis: “*declividade*” e “*distância diagonal da drenagem*”; provavelmente, relacionou-se ao elevado grau de dissecação do relevo.

O mapa de solos obtido com esse modelo foi resultante da aplicação de 212 regras (Tabela 1) e permitiu a identificação de 8 ordens de solos: Argissolos, Latossolos, Neossolos, Gleissolos, Nitossolos, Cambissolos, Espodossolos e Chernossolos. O Argissolo Vermelho-Amarelo (PVA) é o de maior expressão na folha (53%), seguido pelos Neossolos (33%) e Latossolos Amarelos (7%) (Figura 1a). Sete dessas oito ordens apresentaram acurácia individual superior a 50% (Figura 1b). O número final de regras (212) foi obtido depois de se aplicar o método da pré-poda antes do crescimento dos nós da árvore de decisão preservando a quantidade mínima de 10 pixels/polígono de UNIMAP.

Discussão

A partir da interpretação das 212 regras geradas pela árvore de decisão pode-se observar que os Neossolos Litólicos eutróficos (RLe) textura argilosa estão associados a declividades acentuadas, com relevos predominantemente ondulados a montanhosos, com curvaturas em planta geralmente plana e distâncias médias a grandes da drenagem, sendo presente em todas as geologias da folha e independentes da área de contribuição (Figura 1b). Já os RLe textura média são encontrados apenas nas formações Pirambóia com maiores áreas de contribuição. A semelhança dos RL textura média com os de textura argilosa está no fato de serem encontrados em relevos ondulados, de curvatura plana e a distâncias médias da drenagem. Os Neossolos Quartzarênicos também estão associados exclusivamente à geologia Pirambóia na folha de estudo, porém encontram-se presentes em todas as declividades, desde relevos planos a montanhosos, não dependente das curvaturas e geralmente a distâncias médias da drenagem, podendo estar presentes nas outras distâncias também.

Os Cambissolos Háplicos Tb Distróficos estão associados apenas à formação Corumbataí, em áreas de declividade ondulada e distâncias médias da drenagem, independente das curvaturas e da área de contribuição. Os Argissolos Vermelho-Amarelos (PVA) distróficos, textura argilosa, também estão associados a esta geologia, porém, em posições com declividades mais amenas, com relevo predominantemente plano a suave ondulado. A curvatura em planta é geralmente plana, independente do perfil, característica esta presente em todos os solos da geologia Corumbataí. Estes solos encontram-se próximos à drenagem, em distâncias muito baixas a baixas, sendo presentes em todas as áreas de contribuição.

Os Argissolos Vermelho-Amarelos distróficos textura arenosa/média, mesmo grande grupo que o anterior, estão presentes na geologia Corumbataí e Pirambóia. Quando associados à geologia Corumbataí, diferentemente dos argissolos descritos anteriormente, estão presentes em áreas de curvatura convergente ou divergente. Na geologia Pirambóia estão presentes em declividades suaves a forte onduladas e nas diferentes curvaturas em planta, sendo que a curvatura em perfil é geralmente côncava, sendo encontrados geralmente a distâncias médias da drenagem.

Próximos a drenagem e associados exclusivamente aos Sedimentos Aluvionares na folha encontram-se os Gleissolos Háplicos ou Melânicos, presentes em relevos planos, com áreas de contribuição muito baixa ou baixa e distâncias e pequenas da drenagem.

Na formação Itaqueri está presente somente o Latossolo Vermelho-Amarelo textura média, podendo também ser este encontrado em áreas de Sedimentos Aluvionares. Associam-se a declividades menores, com relevo predominantemente plano a suave ondulado, quando em áreas próximas a drenagem. Em distâncias médias, associam-se a relevos mais ondulados. Nessas

duas formações geológicas esses solos estão presentes em todas as curvaturas em planta.

Conclusões

O mapa digital de solos da folha São Pedro, elaborado a partir da mineração de dados geomorfométricos apresentou aproximação com o mapa elaborado por métodos tradicionais, constituindo-se em material adequado para estudos que demandem menor detalhamento.

O conjunto das variáveis geomorfométricas utilizadas permitiu uma acurácia geral de 51% na predição dos solos.

Referências

- [1] SANTANA, D. P. A. 1983. Importância da Classificação dos Solos e do meio ambiente na transferência de tecnologia. Levantamento e Classificação de Solos. Informe Agropecuário, Belo Horizonte.
- [2] McBRATNEY, A.B.; WEBSTER, R. 1981. Spatial dependence and classification of the soil along a transect in northeast Scotland. *Geoderma*, 26: 63-82.
- [3] McBRATNEY, A.B., MENDONÇA-SANTOS, M.L., MINASNY, B. 2003. On digital soil mapping. *Geoderma*, 117: 3-52.
- [4] MENDONÇA-SANTOS, M.L.; SANTOS, H. G.; DART, R. O.; PARES, J. G. 2007. Digital Mapping of Soil Classes, Rio de Janeiro State, Brazil: data, modelling and prediction. In: HARTEMINK, A. E.; McBRATNEY, A. B.; MENDONÇA-SANTOS, M. L. (Org.). *Digital Soil Mapping with Limited Data*. Amsterdam: Elsevier.
- [5] LAGACHERIE, P.; McBRATNEY, A. B. 2007. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: *Digital Soil Mapping: An Introductory Perspective*. Montpellier and Sydney, p. 1-3.
- [6] HAN, J.; KAMBER, M. 2001. *Data Mining - Concepts and Techniques*. 1a ed. Nova York: Morgan Kaufmann.
- [7] BREIMAN, L.; FRIEDMAN, J.H.; OLSEN, R.A. 1984. *Stone, Classification and Regression Trees*. Wadsworth.
- [8] IPT (Instituto de Pesquisas Tecnológicas do Estado de São Paulo). 1981. *Mapa geomorfológico do estado de São Paulo*. Vol. 1: Nota Explicativa e Vol. 2: Mapa. São Paulo: IPT, (Série Monografias, 5).
- [9] ALMEIDA, F.F.M. 1964. *Fundamentos Geológicos do Relevo Paulista*. 99p. Tese (Doutorado). Universidade de São Paulo, Instituto de Geografia, São Paulo.
- [10] EMBRAPA. 2006. *Sistema Brasileiro de Classificação de Solos*. Rio de Janeiro: Embrapa Solos. 2 ed. 316p.
- [11] ESRI 2007. *GIS and Mapping Software*. Homepage: <http://www.esri.com>.
- [12] LOPES-ASSAD, M.L. 1995. Uso de sistema de informações geográficas na determinação da aptidão agrícola de terras. *Revista Brasileira de Ciência do Solo*, 19: 133-139.
- [13] VALERIANO, M.M. 2003. Curvatura vertical de vertentes em microbacias pela análise de modelos digitais de elevação. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 7: 539-546.
- [14] ITC. 2001. *Ilwis 3.3: User's guide*. Enschede, ITC, 530p.
- [15] CLARK LABS. 2006. *Idrisi Andes*. Clark Labs, Clark University, Worcester, MA, USA.
- [16] WEKA. 2006. *Data Mining Software in Java*. University of Waikato. Homepage: <http://www.cs.waikato.ac.nz/ml/weka>.

Tabela 1. Número de regras geradas pela árvore de decisão e acurácia geral do modelo nos diferentes balanceamentos de classes.

	Balanceamento de classes		
	0	0,5	1
Regras	212	362	362
Acurácia geral do modelo (%)	54,3	49,3	15,1
Acurácia geral retiradas unidades com probabilidade zero (%)	54,8	50,5	33,4

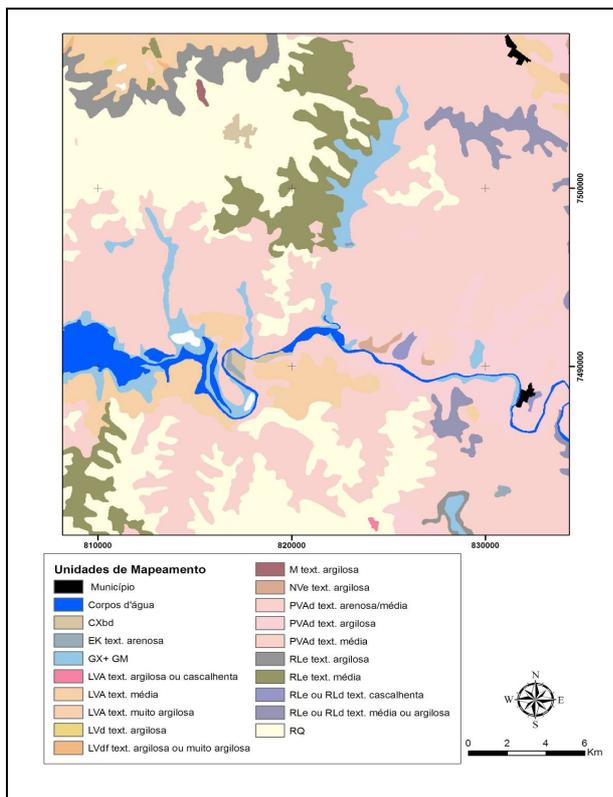
Tabela 2. Acurácia de todas as unidades de mapeamento de solos da folha São Pedro

Unidades de Mapeamento	Balanceamento de classes		
	0	0,5	1
	-----%-----		
1) LVA textura média	66,1	65,3	56,4
2) LVA textura muito argilosa	0	11,7	8,2
3) RLe textura argilosa	77,6	75	55,2
4) RQ	49,4	46,6	35,8
5) PVAd textura arenosa/média	53,6	54,5	65,6
6) PVAd textura argilosa	49,4	46,9	53,6
7) NVe textura argilosa	0	4,7	2,3
8) LVdf textura argilosa ou muito argilosa	0	4,8	3,6
9) RLe textura média	11,1	31,6	16,9
10) LVd textura argilosa	0	4,6	4,6
11) PVAd textura média	0	6,1	4,1
12) M textura argilosa	0	1,1	0,3
13) GX + GM	64,3	46,9	30,2
14) RLe ou RLd textura média ou argilosa	57,1	17,8	16,2
15) CXbd	75	24	1,7
16) RLe ou RLd textura cascalhenta	0	3	2,4
17) EK text. arenosa	0	5,7	1,6
18) LVA text. argilosa ou cascalhenta	0	0	0,1

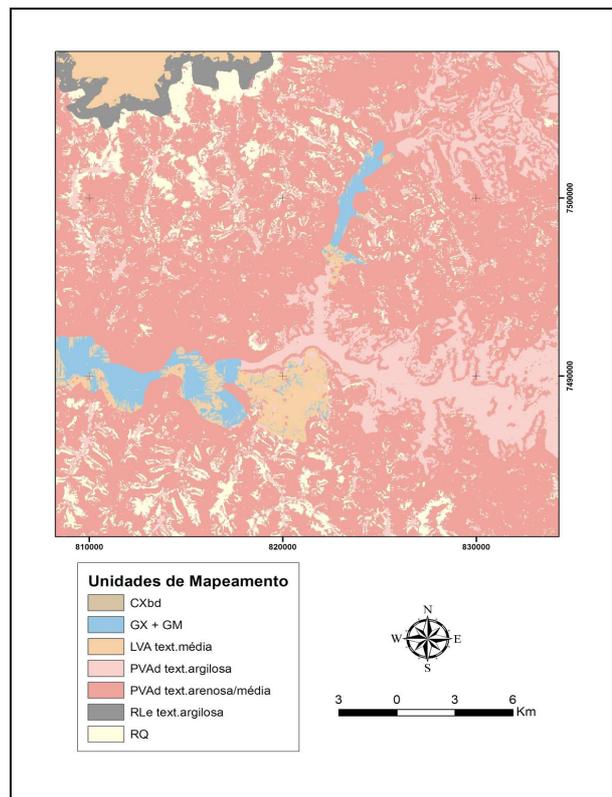
1 - Latossolo Vermelho-Amarelo textura média; 2 - Latossolo Vermelho-Amarelo textura muito argilosa; 3 - Neossolo Litólico eutrófico textura argilosa; 4 - Neossolo Quartzarênico; 5 - Argissolo Vermelho Amarelo distrófico textura arenosa sob média; 6 - Argissolo Vermelho Amarelo distrófico textura argilosa; 7 - Nitossolo Vermelho eutrófico textura argilosa; 8 - Latossolo Vermelho distrófico textura argilosa ou muito argilosa; 9 - Neossolo Litólico eutrófico textura média; 10 - Latossolo Vermelho distrófico textura argilosa; 11 - Argissolo Vermelho-Amarelo distrófico textura média; 12 - Chernossolo textura argilosa; 13 - Gleissolo Háptico + Melânico; 14 - Neossolo Litólico eutrófico ou distrófico cascalhento; 15 - Cambissolo Háptico distrófico; 16 - Neossolo Litólico eutrófico ou distrófico textura média ou média sob argilosa; 17 - Espodossolo Humilúvico textura arenosa; 18 - Latossolo Vermelho Amarelo textura argilosa ou cascalhenta

Tabela 3. Acurácia das unidades de mapeamento individualmente, retiradas as unidades de solos com probabilidade nula de ocorrência no modelo gerado para os dados brutos

Unidades de Mapeamento de Solos	Balanceamentos		
	0	0,5	1
LVA textura média	68,9	72,4	63,7
RLe textura argilosa	79,1	62,2	50,9
RQ	49	47	35,8
PVAd textura arenosa/media	53,9	55,6	60,9
PVAd textura argilosa	51,1	48,3	50,8
RLe textura média	16,7	25,5	16
GX + GM	65,5	38,4	32,1
RLe ou RLd textura média ou argilosa	0	14,9	14
CXbd	71,4	9,7	1,8



(a)



(b)

Figura X. Mapa de solos da Folha de São Pedro após unificação e atualização da legenda (a) e produzido a partir de variáveis geomorfométricas (b)