



Mapeamento pedológico digital por meio de algoritmos de mineração de dados aplicados a diferentes conjuntos de variáveis preditivas.⁽¹⁾

Laura Milani da Silva Dias⁽²⁾; Ricardo Marques Coelho⁽³⁾; Gustavo Souza Valladares⁽⁴⁾.

⁽¹⁾ Trabalho executado com recursos Petrobrás (Rede de Sistemas de Produção/PBIO).

⁽²⁾ Mestre; Instituto Agronômico; Campinas, São Paulo; laurads5@yahoo.com.br; ⁽³⁾ Pesquisador; Instituto Agronômico; ⁽⁴⁾ Professor Doutor; Universidade Federal do Piauí.

RESUMO: Em área da bacia sedimentar do São Francisco foi realizado o mapeamento pedológico digital por meio da predição de algoritmos de mineração de dados aplicados a diferentes conjuntos de variáveis do relevo, geologia e índices derivados de sensor remoto orbital. O objetivo foi avaliar quais conjuntos de variáveis geravam mapas mais acurados. Para tanto, foram elaboradas quatro matrizes de dados cujas classes de solo a serem preditas foram obtidas em uma amostragem de solo no campo com classificação em dois níveis taxonômicos. Os resultados apontaram melhor desempenho para as matrizes que consideram relevo e geologia como variáveis preditivas quando comparadas às matrizes compostas por geologia e índices derivados de sensor remoto orbital. A superioridade ocorre tanto para os mapas com classes de solo em quarto nível categórico e grupamento textural quanto para aqueles com classificação em segundo nível categórico.

Termos de indexação: mapeamento digital de solos, modelos preditivos, acurácia.

INTRODUÇÃO

A execução de mapeamentos de solo é uma demanda permanente por ser uma importante ferramenta para o uso e ocupação racional das terras (Mendonça-Santos & Santos, 2003). Apesar disso, os recursos financeiros destinados aos mapeamentos de solo no Brasil vêm diminuindo muito, bem como o número de profissionais capacitados a realizar essa tarefa (Chagas et al., 2010).

Novas possibilidades tecnológicas vêm sendo investigadas para o levantamento de solos com o objetivo de produzir mapas digitais que diminuam a necessidade de trabalhos de campo e análise laboratorial posterior (McBratney et al., 2003).

Como principal aplicação dessa abordagem está a predição das classes e propriedades de solos por meio de modelos matemáticos e seu mapeamento digital de forma contínua e espacial (McBratney et al., 2003).

O mapeamento digital de solos (MDS) tem experimentado um rápido desenvolvimento de

métodos devido principalmente à crescente disponibilidade de mapas auxiliares, destacando o uso das imagens de sensoriamento remoto orbital e os atributos do terreno derivados de modelos digitais de elevação (Dobos et al., 2000).

Ainda não há um conhecimento sistematizado sobre a variabilidade espacial do solo e, em especial, para o MDS existem diversas possibilidades metodológicas a serem analisadas.

O objetivo do estudo foi produzir mapas pedológicos digitais por meio de técnicas de mineração de dados aplicadas a dados do meio físico, de sensor remoto orbital e pontos de observação de solo no campo e validá-los a fim de avaliar qual o conjunto de variáveis preditivas gera mapas mais confiáveis.

MATERIAL E MÉTODOS

A área estudada está inserida na microrregião de Montes Claros no norte do Estado de Minas Gerais.

De acordo com a classificação de Köppen, o clima da região é do tipo Aw, tropical de savana com estação seca de inverno e médias anuais para precipitação e temperatura de 1.060 mm e 24,20 °C respectivamente (PMMC, 2006).

A partir da compilação do mapa geológico em escala 1.100.000 (CPRM, 2011) e do mapa metalogenético em escala 1:250.000 (CPRM, 1985), ambos da folha Montes Claros foram reconhecidas as formações geológicas: 1 calcilitos e calcarenitos, 2 depósitos aluvionares, 3 depósitos colúvio-eluviais, 4 metassilitos e folhelhos e 5 ritmitos.

Geomorfologicamente a área encontra-se na Bacia Sedimentar do Rio São Francisco com o predomínio de superfícies de aplainamento e planaltos residuais (Leite & Brito, 2012).

A partir de dados de hipsometria SRTM com resolução espacial de 90 metros foi gerado um modelo digital de elevação (MDE) e derivadas as variáveis de relevo altitude, declividade, curvatura planar, curvatura em perfil, orientação das vertentes, distância da drenagem e índice topográfico de umidade.

Uma imagem do sensor OLI do Landsat 8 foi utilizada para derivar os índices "Clay minerals -



CMI”, “Iron Oxide - IOI” e “Normalized difference vegetation index – NDVI.

Para obtenção das informações dos solos, as variáveis geradas foram inseridas no programa de amostragem aleatório estratificada (cLHS) como condicionantes à seleção de 300 pontos (Wyss & Jorgensen, 1998).

O solo foi observado por sondagem com trado até 1 metro de profundidade ou em cortes de estradas em que foi caracterizado morfologicamente e coletadas amostras nos horizontes de superfície (horizonte A) e subsuperfície (horizonte B) de cerca de 1/3 do total de pontos para análises laboratoriais. Realizaram-se análises granulométricas e químicas para rotina pedológica e classificação por meio do Sistema Brasileiro de Classificação de Solos até o quarto nível categórico e grupamento textural (EMBRAPA, 2013).

Do total de 300 observações de solo classificado, foram selecionadas 252 que representam os solos de maior ocorrência na área divididas em 190 para treinamento dos modelos preditivos e 62 para posterior validação dos mapas.

As matrizes de dados foram elaboradas com duas combinações entre as variáveis sendo a primeira com as variáveis preditivas de relevo e geologia e a segunda com geologia e os índices derivados da imagem Landsat 8 para prever classes de solo em dois níveis taxonômicos (Tabela 1).

Cada matriz tem 136.149 linhas, destas, apenas 190 possuem classe de solo (observações de solo em campo) resultando em uma proporção de 0,14%, as demais deverão ser classificadas pelos modelos preditivos.

Tabela 1: Descrição das matrizes de dados.

Matriz	Nível taxonômico	Variáveis preditivas
1	4º nível + textura	Geologia + Relevo
2	4º nível + textura	Geologia + Índices Landsat 8
3	2º nível	Geologia + Relevo
4	2º nível	Geologia + Índices Landsat 8

O treinamento dos modelos foi realizado no software de mineração de dados Weka 3.6.6 (Witten & Franck, 2005). Foi verificado o desempenho de três algoritmos sendo J48 (árvores de decisão) *Random Forest* (combinação de diversas árvores) e MLP (redes neurais) associados a técnicas de pré-processamento que buscam elevar a acurácia como discretização, balanceamento de classes, ordenamento e seleção de variáveis. Desta maneira, o melhor modelo preditivo por matriz de dados foi

selecionado e aplicado para predição das classes de solo faltantes e posteriormente foram gerados e validados os mapas pedológicos digitais.

A validação de cada um dos mapas foi realizada por meio dos 62 pontos classificados em campo. Esses dados foram confrontados com os mapas preditos por meio de matrizes de erro e a partir desses resultados foram calculadas as medidas de acurácia exatidão global e índice kappa (Landis & Koch, 1997).

RESULTADOS E DISCUSSÃO

O mapa pedológico digital gerado a partir da matriz de dados 1 (Figura 1a) apresenta como classes de maior expressão os Cambissolos Háplicos Tb distróficos típicos com textura argilosa (21,2%) e os Latossolos Vermelhos distróficos típicos de textura argilosa (20,6%) bem distribuídos por toda a extensão do mapa.

Os Latossolos Vermelhos distróficos típicos com textura média e os Latossolos Vermelhos eutróficos que se concentram no norte da área também representam juntos área significativa (20,8%).

As classes preditas que ocuparam menor área foram os Latossolos Vermelho-Amarelos distróficos típicos, textura média (0,8%) e argilosa (2,6%), os Argissolos Vermelhos (1,9%) e os Neossolos Regolíticos (2%) e Litólitos (2,3%) eutróficos típicos de textura média.

A exatidão global do mapa foi de 53,2% e o índice Kappa de 0,48 (concordância moderada) (Tabela 2).

Tabela 2: Valores para exatidão global (%), índice Kappa e grau de concordância por mapa validado.

Matriz	Exatidão global	Índice Kappa	
		Valor	Concordância
1	53,2	0,48	Moderada
2	27,0	0,20	Ligeira
3	54,8	0,41	Moderada
4	32,3	0,13	Ligeira

O mapa gerado para a matriz de dados 2 obteve acurácia inferior (27%) quando comparado ao mapa da matriz de dados 1 e o índice kappa indica apenas uma concordância ligeira entre as classes de solo reais e as que foram preditas (0,20) (Tabela 2). Este resultado se deve ao não uso das variáveis do relevo para prever as classes de solo (Figura 1b).

A área ocupada pelos Cambissolos Háplicos Tb distróficos típicos (13,6%) diminui significativamente e neste mapa se concentram no norte da área. Os



Latossolos Vermelhos distróficos típicos com textura argilosa também diminuem sua área relativa (14,8%), enquanto classes como os Nitossolos Vermelhos distróficos (4,3%) e eutróficos (9,5%) aumentam. Além disso, este mapa apresenta uma grande concentração de Argissolos Vermelhos no extremo sul da área.

O mapa digital gerado a partir da matriz de dados 3 considera as classificações de solo em segundo nível categórico e, por este motivo, é menos fragmentado que os anteriores (Figura 1c).

Neste mapa, os Cambissolos representam 26% da área predita e se concentram na parte inferior junto aos Argissolos Vermelhos (3,5%) e Vermelho-Amarelos (4%), enquanto os Latossolos Vermelhos representam 46% da área.

A validação deste mapa mostra exatidão global de 54,83% e índice Kappa 0,41 (concordância moderada) (Tabela 2).

No mapa gerado a partir da matriz de dados 4 (Figura 1d), sem as variáveis do relevo como preditivas das classes de solo em segundo nível categórico, a área dos Latossolos Vermelhos diminui significativamente (22,1%). Em contrapartida, houveram aumentos nas áreas preditas como Cambissolos Háplicos (40,4%), Argissolos Vermelhos (5,6%), Neossolos Litólicos (4,7%) e Regolíticos (3,8%). Sua exatidão global é de 32,25% e índice Kappa 0,13 que indica apenas uma concordância ligeira entre os dados reais e os dados preditos (Tabela 2).

CONCLUSÕES

Os mapas que utilizam as variáveis preditivas geologia e relevo apresentam maior acurácia tanto com classes de solo em quarto nível categórico e agrupamento textural quanto em segundo nível categórico. Esta superioridade com relação ao uso da geologia associada a índices derivados de imagem Landsat 8 sugere controle litológico e morfométrico da distribuição dos solos na área de estudo.

REFERÊNCIAS

CHAGAS, C. da S. FERNANDES FILHO, E. I. VIEIRA, C. A. O. SCHAEFER, C. E. G. R. CARVALHO JUNIOR, W de. Atributos topográficos e dados do Landsat 7 no mapeamento digital de solos com uso de redes neurais. Pesquisa Agropecuária Brasileira, Brasília, v. 45, n. 5, p. 497-507, maio 2010.

DOBOS, E.; MICHELI, E.; BAUMGARDNER, M.F.; BIEHL, L.; HELT, T. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma, v.97, p.367-391, 2000.

EMBRAPA. Sistema Brasileiro de Classificação de Solos. 3.ed. Brasília: Embrapa Produção de Informação; Rio de Janeiro: Embrapa Solos, 2013. 353p

LANDIS, J. R. & KOCH, G. G. The measurement of observer agreement for categorical data. Biometrics, v.33, p.159-174, 1977.

LEITE, M.R.; BRITO, J.L.S. Mapeamento morfoestrutural e morfoescultural na região de cerrado no Norte de Minas Gerais. Soc. & Nat., Uberlândia, ano 24 n. 1, 115-126, jan/abr. 2012

MAPA GEOLÓGICO. Folha Montes Claros - SE-23-X-A-VI. Escala 1.100.000. CPRM, 2011

MAPA METALOGENÉTICO. Folha Montes Claros - SE-23-X-A. Escala 1: 250.00. CPRM, 1985.

McBRATNEY, A.B.; SANTOS, M.L.M.; MINASNY, B. On digital soil mapping. Geoderma, v.117, p.3-52, 2003.

MENDONÇA-SANTOS, M. L.; SANTOS, H. G. dos. Mapeamento digital de solos e atributos de solos: métodos, paradigmas e novas técnicas. Rio de Janeiro: Embrapa Solos, 2003.

PMMC, Prefeitura Municipal de Montes Claros, 2006. Coletânea de Informações sobre o Município de Montes Claros, 2006. Disponível em: <http://www.pmmc.br>, acesso em 18/05/2015.

SABINS, F. F. Remote Sensing: Principles and Interpretation. 3rd ed. New York: W. H. Freeman and Company, 1997. 432p.

WITTEN, I.H., FRANK, E. Data mining: practical machine learning tools and techniques. 2nd edition. San Francisco: Morgan Kaufmann; 2005.

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)

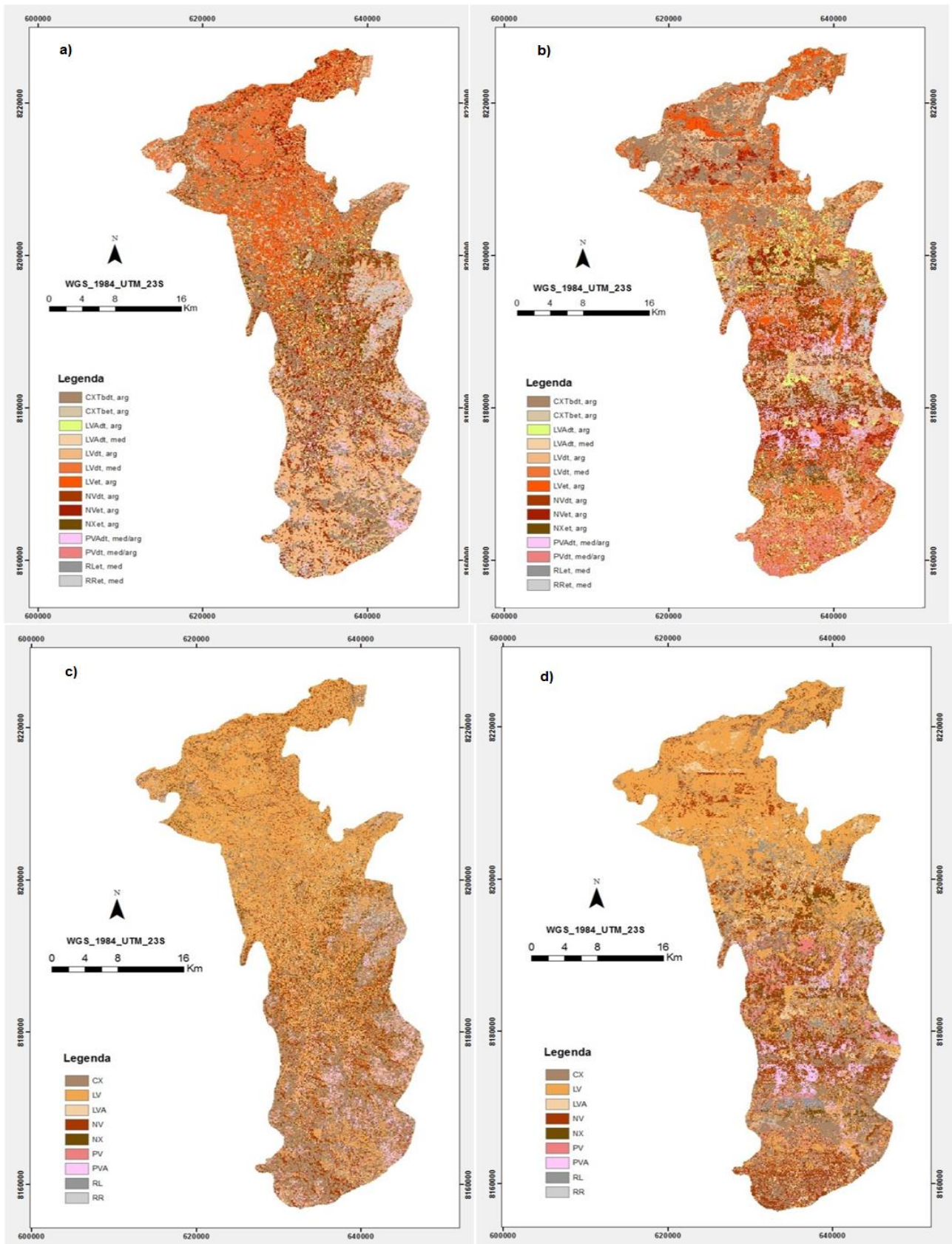


Figura 1: Mapas pedológicos digitais gerados por meio de algoritmos de mineração de dados aplicados a diferentes conjuntos de variáveis preditivas: (a) Mapa derivado da matriz de dados 1, (b) Mapa derivado da matriz de dados 2, (c) Mapa derivado da matriz de dados 3 e (d) Mapa derivado da matriz de dados 4.