



## Predição Espacial da Textura Superficial do Solo por Árvore de Regressão e Redes Neurais Artificiais

**Cesar da Silva Chagas<sup>(1)</sup>; Waldir de Carvalho Junior<sup>(1)</sup>; Silvio Barge Bhering<sup>(1)</sup>; Nilson Rendeiro Pereira<sup>(2)</sup>.**

<sup>(1)</sup> Pesquisador A; Empresa Brasileira de Pesquisa Agropecuária (Embrapa Solos); Rio de Janeiro, RJ; cesar.chagas@embrapa.br; waldir.carvalho@embrapa.br; silvio.bhering@embrapa.br; <sup>(2)</sup> Pesquisador B; Empresa Brasileira de Pesquisa Agropecuária (Embrapa Solos); Rio de Janeiro, RJ; nilson.pereira@embrapa.br.

**RESUMO:** A textura é uma propriedade física do solo importante e altamente variável que influencia grandemente muitas outras propriedades de grande importância para a produção agrícola como a fertilidade e a capacidade de retenção de umidade. O objetivo desse estudo foi avaliar a eficiência de dados do sensor TM do Landsat 5 no mapeamento digital de solos no semiárido brasileiro por meio da utilização de árvores de regressão (AR) e redes neurais artificiais (RNA), ambas implementadas no software livre R. Foram utilizadas na predição da areia, silte e argila, 399 amostras da camada superficial (0 - 20 cm) dos solos e como covariáveis as bandas 1, 2, 3, 4, 5 e 7, o índice NDVI e as relações entre as bandas 3 e 2, bandas 3 e 7 e bandas 5 e 7. Os resultados das análises realizadas, usando um conjunto de validação independente, mostraram que as melhores estimativas dos atributos foram obtidas com a utilização dos modelos RNA, que explicou 57% da variabilidade espacial da areia, 49% da argila e 32% do silte. A RNA mostrou-se mais vantajosa, em comparação com a AR, pelo fato de não ser sensível ao sobreajustamento (overfitting) e nem a presença de ruídos nos dados. Além disso, a RNA produziu mapas da distribuição dos atributos mais realísticos do que a AR.

**Termos de indexação:** mapeamento digital de solos, variabilidade espacial, pedometria.

### INTRODUÇÃO

A informação sobre solos, incluindo a variabilidade de seus atributos, é fundamental para a formulação de políticas agrícolas, manejo dos solos e para o monitoramento dos impactos ambientais advindos de seu uso.

Para Boettinger et al. (2008), dados de sensores remotos orbitais podem ser utilizados como covariáveis ambientais no mapeamento digital dos solos, especialmente em regiões áridas e semiáridas, facilitando, dessa maneira, o mapeamento e reduzindo a necessidade de levantamentos de campo onerosos e demorados (Mulder et al., 2011). Vários estudos têm demonstrado as relações entre diferentes atributos

do solo e dados de sensores remotos (Demattê et al., 2007; Breunig et al., 2008; Liao et al., 2013), sendo que os métodos mais comumente utilizados para a predição desses atributos têm sido a regressão linear múltipla, a regressão por mínimos quadrados parciais e os métodos geoestatísticos. A aplicação de métodos de mineração de dados como RNA (Zhao et al., 2009; Viscarra Rossel & Behrens, 2010) e AR (Henderson et al., 2005; Ließ et al., 2012) tem sido menos comum.

O presente estudo tem por objetivo comparar a eficiência de modelos de AR e de RNA na predição da textura superficial dos solos, utilizando dados do sensor TM do Landsat 5 como covariáveis ambientais em uma área do semiárido brasileiro.

### MATERIAL E MÉTODOS

#### Localização e meio físico da área de estudo

O estudo foi realizado em parte da área pertencente ao projeto de irrigação Salitre, no município de Juazeiro, Estado da Bahia. A área possui aproximadamente 35000 ha e está localizada entre as coordenadas UTM 8.909.500 e 8.938.300m N e 324.300 e 348.800m E, zona 24S. O clima na região é do tipo BShw' (clima semiárido com inverno seco e verão chuvoso), segundo Köppen, e como consequência a vegetação original é formada pela caatinga hiperxerófila. O relevo é essencialmente plano e a geologia formada por calcários da Formação Caatinga e por rochas gnáissicas-graníticas do complexo Caraíba-Paramirim (Souza et al., 2003). As classes de solo mais representativas são os Vertissolos, Cambissolos e Planossolos.

#### Atributos do solo e covariáveis ambientais

Para predição da areia, silte e argila foram utilizados dados da camada superficial (0 - 20 cm) de 399 perfis de solos disponibilizados pela CODEVASF, coletados no ano de 1989. As frações da composição granulométrica foram determinadas conforme Embrapa (1979).

Foram utilizados como covariáveis dados do sensor TM do Landsat 5 (números digitais), órbita/ponto 217/67, com resolução espacial de 30m, de janeiro de 2007, conforme se segue: banda 1 (0,450-0,515  $\mu\text{m}$ ), banda 2 (0,525-0,605  $\mu\text{m}$ ),



banda 3 (0,630-0,690  $\mu\text{m}$ ), banda 4 (0,755-0,900  $\mu\text{m}$ ), banda 5 (1,550-1,750  $\mu\text{m}$ ), banda 7 (2,090-2,350  $\mu\text{m}$ ), índice NDVI (banda 4 - banda 3/banda 4 + banda 3) e as relações entre a banda 3 e a banda 2 (b3/b2), entre a banda 3 e a banda 7 (b3/b7) e entre a banda 5 e a banda 7 (b5/b7).

### Modelos de predição

Neste estudo foram utilizadas AR e RNA. AR é um método hierárquico que utiliza o particionamento recursivo binário, no qual os dados da variável dependente são divididos sucessivamente com base na importância das variáveis independentes, em dois nós, sendo esse processo repetido de maneira recursiva até que os subgrupos alcancem um tamanho mínimo ou até que nenhuma melhoria possa ser alcançada (Breiman et al., 1984). Os parâmetros de configuração utilizados foram: método Anova (regra para subdivisão dos dados) e *cp* (parâmetro de complexidade) inicial de 0,01 (default).

A RNA é uma técnica de inteligência artificial que tenta construir um modelo matemático que supostamente trabalha de maneira análoga ao cérebro humano. Inicialmente, para sua utilização os valores das variáveis e covariáveis foram reescalados entre 0 e 1 para facilitar o processo de treinamento e evitar a saturação da rede. Na fase de treinamento diferentes arquiteturas foram testadas (5 a 30 neurônios na camada interna), sempre utilizando o algoritmo de aprendizado *backpropagation*, com alocação aleatória dos pesos interneurônios entre -0,5 e 0,5 e uma taxa de aprendizado de 0,2, considerando 1.000 ciclos de aprendizagem.

As ARs e RNAs foram implementadas no software R (R Development Core Team, 2013), através das funções *rpart* (AR) e do pacote *RNNNS* (RNA).

A performance dos modelos foi medida, usando amostras de validação independentes (80 amostras que correspondem a 20% do conjunto total), não utilizadas na fase de treinamento (319 amostras), pelo cálculo da correlação entre os valores observados e os valores estimados através do coeficiente de determinação ( $R^2$ ).

## RESULTADOS E DISCUSSÃO

Inicialmente, a análise dos modelos de AR mostrou que as covariáveis tiveram diferentes participações e importância para a predição dos atributos avaliados, conforme **tabela 1**. Para todas as frações a covariável mais importante foi a relação B3/B7, seguida da B5 (areia e argila) e B3 (silte). Por outro lado, as relações entre as bandas B3/B2 e

B5/B7 mostraram-se pouco importantes. Os solos da área são formados predominantemente por minerais de argila 2:1 e assim, a pouca variabilidade com relação ao tipo de mineral de argila pode explicar a ausência de correlação significativa entre os atributos e a covariável B5/B7.

**Tabela 1** - Importância das covariáveis estimada pelas ARs.

Atributo	Covariáveis
Areia	B3/B7 > B5 > B2 > B3 > NDVI > B1 > B7 > B4 > B3/B2 > B5/B7
Silte	B3/B7 > B3 > B1 > B5 > B2 > B7 > NDVI > B3/B2 > B5/B7 > B4
Argila	B3/B7 > B5 > B2 > B3 > B1 > NDVI > B7 > B4 > B5/B7 > B3/B2

Com relação as RNAs, dentre as arquiteturas de redes treinadas, o melhor desempenho, medido pela soma do erro ao quadrado (SSE, sigla em inglês), foi obtido pela RNA com 10 neurônios na camada interna para a areia, com 12 neurônios para o silte e com 10 neurônios para a argila. Sendo assim, estas arquiteturas foram escolhidas para a predição dos atributos avaliados.

Os resultados obtidos pelos modelos AR e RNA, a partir da utilização do conjunto de amostras de validação independente, são apresentados na **tabela 2**.

**Tabela 2** - Resultados das predições dos modelos utilizados.

Atributo	$R^2$		
	Árvore de Regressão	Redes Neurais Artificiais	Diferença
Areia	0,48	0,57	0,09
Silte	0,13	0,32	0,19
Argila	0,42	0,49	0,07

Os resultados obtidos pelos modelos de AR podem ser considerados como moderadamente satisfatórios para a areia e argila e insatisfatório para o silte. Henderson et al. (2005) utilizaram AR, atributos do terreno e dados derivados do Landsat, como covariáveis e obtiveram resultados similares aos deste estudo na predição da argila na camada superficial dos solos ( $R^2 = 0,44$ ).

Lamsal & Mishra (2010) utilizaram modelos de AR juntamente com atributos do terreno, informações geológicas e fisiográficas no mapeamento das frações texturais do solo em uma microbacia na Florida e obtiveram resultados para a camada superficial dos solos ligeiramente inferiores para a areia e argila (40 e 38%, respectivamente) e ligeiramente superior para o silte (25%) aos obtidos neste estudo, que variaram de 13 a 48% (**Tabela 2**).



Ließ et al. (2012) utilizaram atributos do terreno derivados de um modelo digital de elevação, combinados com dois métodos de determinação da composição granulométrica (pipeta e laser) e obtiveram resultados inferiores aos encontrados no presente estudo para areia (0,03 a 15%), silte (0,3 a 11%) e argila (29 a 43%) na camada superficial dos solos. A baixa performance dos modelos AR neste estudo foi atribuída ao tamanho reduzido do conjunto de dados utilizado.

Os resultados obtidos pelos modelos de RNA também foram considerados como moderadamente satisfatórios para a areia e argila e insatisfatório para o silte. Viscarra Rossel & Behrens (2010), utilizaram métodos de mineração de dados, entre eles RNA, para modelar os teores de argila no solo, a partir de dados de reflectância difusa no visível e infravermelho próximo obtidos em laboratório e encontraram resultados de  $R^2$  (0,86 a 0,88) bem superiores aos encontrados no presente estudo, que utilizou dados de um sensor remoto orbital.

Zhao et al. (2009) desenvolveram um modelo de RNA para predição de areia e argila baseado na utilização de atributos derivados de um mapa de solos de pequena escala e atributos do terreno derivados de um modelo digital de elevação (mde). A acurácia relativa geral (ROA) variou de 81% para areia a 88% para argila, superiores portanto, aos deste estudo.

O modelo RNA foi superior ao modelo AR na predição de todos os atributos avaliados (**Tabela 2**), sendo que a maior diferença foi verificada para o silte (0,19). A comparação do desempenho de modelos AR e RNA não foi encontrada em nenhum estudo.

De acordo com Wetterlind & Stenberg (2010), a baixa performance dos modelos de RNA e AR na predição do silte pode estar relacionada com as condições específicas da área de estudo, aliada a maior dificuldade de se distinguir esta fração, que pode ser assumida como sendo uma mistura de areia e argila, nas regiões do visível e do infravermelho.

Demattê et al. (2007) ressalta que a quantificação de atributos do solo a partir de um sensor orbital não é uma tarefa simples, devido à complexidade dos solos. Neste sentido, resultados de  $R^2$  próximos de 0,5, como os obtidos para areia e argila pelo modelo RNA, no presente estudo, provavelmente são devidos à interferência física destes constituintes do solo na energia incidente e refletida.

Os teores das frações avaliadas variaram, nas amostras de treinamento de 116 a 790 g.kg<sup>-1</sup> (areia), 54 a 344 g.kg<sup>-1</sup> (silte) e de 133 a 636 g.kg<sup>-1</sup> (argila). Assim, o modelo RNA produziu valores de distribuição espacial desses atributos mais próximos

dos valores originais do que os modelos AR (**Figura 1**). Os teores de areia variaram de 211 a 541 g.kg<sup>-1</sup>, no modelo AR e de 133 a 915 g.kg<sup>-1</sup>, no modelo RNA. Com relação ao silte, os teores variaram entre 146 e 265 g.kg<sup>-1</sup> (AR) e entre 32 e 338 g.kg<sup>-1</sup> (RNA). Por sua vez, a argila variou de 285 a 543 g.kg<sup>-1</sup>, no modelo AR e no modelo RNA de 102 a 620 g.kg<sup>-1</sup>.

## CONCLUSÕES

O modelo RNA apresentou coeficientes de determinação ( $R^2$ ), considerando a amostras de validação, superiores aos obtidos pelo modelo AR na predição de todas as frações avaliadas.

O uso de covariáveis obtidas por sensoriamento remoto orbital aliado a abordagem por redes neurais mostrou potencial de utilização para estimar valores de granulometria, em especial areia e argila, para condições de solos de clima semiárido.

## AGRADECIMENTOS

Os autores agradecem a CODEVASF, representada por Luiz Augusto Costa Fernandes, pela cessão dos dados que permitiram a realização deste estudo.

## REFERÊNCIAS

- BOETTINGER, J. L. et al. Landsat Spectral Data for Digital Soil Mapping. In: HARTEMINK, A. E. et al. (Eds.). Digital Soil Mapping with limited data. New York, Springer-Verlag, 2008. p.192-202.
- BREIMANN, L. et al. Classification and Regression Trees. CRC press, Wadsworth. 1984.
- BREUNIG, F. M.; GALVÃO, L. S. & FORMAGGIO, A. R. Detection of sandy soil surfaces using ASTER-derived reflectance, emissivity and elevation data: potential for the identification of land degradation. Int. J. Remote Sens. v.29, p.1833-1840, 2008.
- DEMATTÊ, J. A. M. et al. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. Int. J. Remote Sens., v.28, p.3813-3829, 2007.
- EMBRAPA. Serviço Nacional de Levantamento e Conservação de Solos (Rio de Janeiro, RJ). Manual de métodos de análise de solos. Rio de Janeiro, 1979.
- Henderson, B.L. et al. Australia-wide predictions of soil properties using decision trees. Geoderma, v.124, p.383-398, 2005.
- LAMSAL, S.; MISHRA, U. Mapping soil textural fractions across a large watershed in northeast Florida. Journal of environmental management, v. 91, p.1686-1694, 2010.

LIAO, K. et al. Spatial estimation of surface soil texture using remote sensing data. *Soil Science and Plant Nutrition*, v.59, p.488-500, 2013.

LIEß, M.; GLASER, B.; HUWE, B. Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models. *Geoderma*, v.170, p.70-79, 2012.

MULDER, V. L. et al. The use of remote sensing in soil and terrain mapping - A review. *Geoderma*, v.162, 1-19, 2011.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2013. Disponível em <<http://www.r-project.org/isbn-3-900051-07-0>>. Acesso em 08 de maio de 2013.

SOUZA, D. J. et al. Mapa geológico do Estado da Bahia - escala 1:1. 000.000. Salvador: CPRM, 2003. Versão 1.1. Programas Carta Geológica do Brasil a milionésimo e Levantamentos geológicos básicos do Brasil (PLGB).

VISCARA ROSSEL, R. A. & BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, v.158, p.46-54, 2010.

WETTLERLIND, J. & STENBERG, B. Near-infrared spectroscopy for within-field soil characterization: Small local calibrations compared with national libraries spiked with local samples. *Eur. J. Soil Sci.*, v. 61, p.823-843, 2010.

ZHAO, Z. et al. Predict soil texture distributions using an artificial neural network model. *Computers and Electronics in Agriculture*, v. 65, p.36-48, 2009.

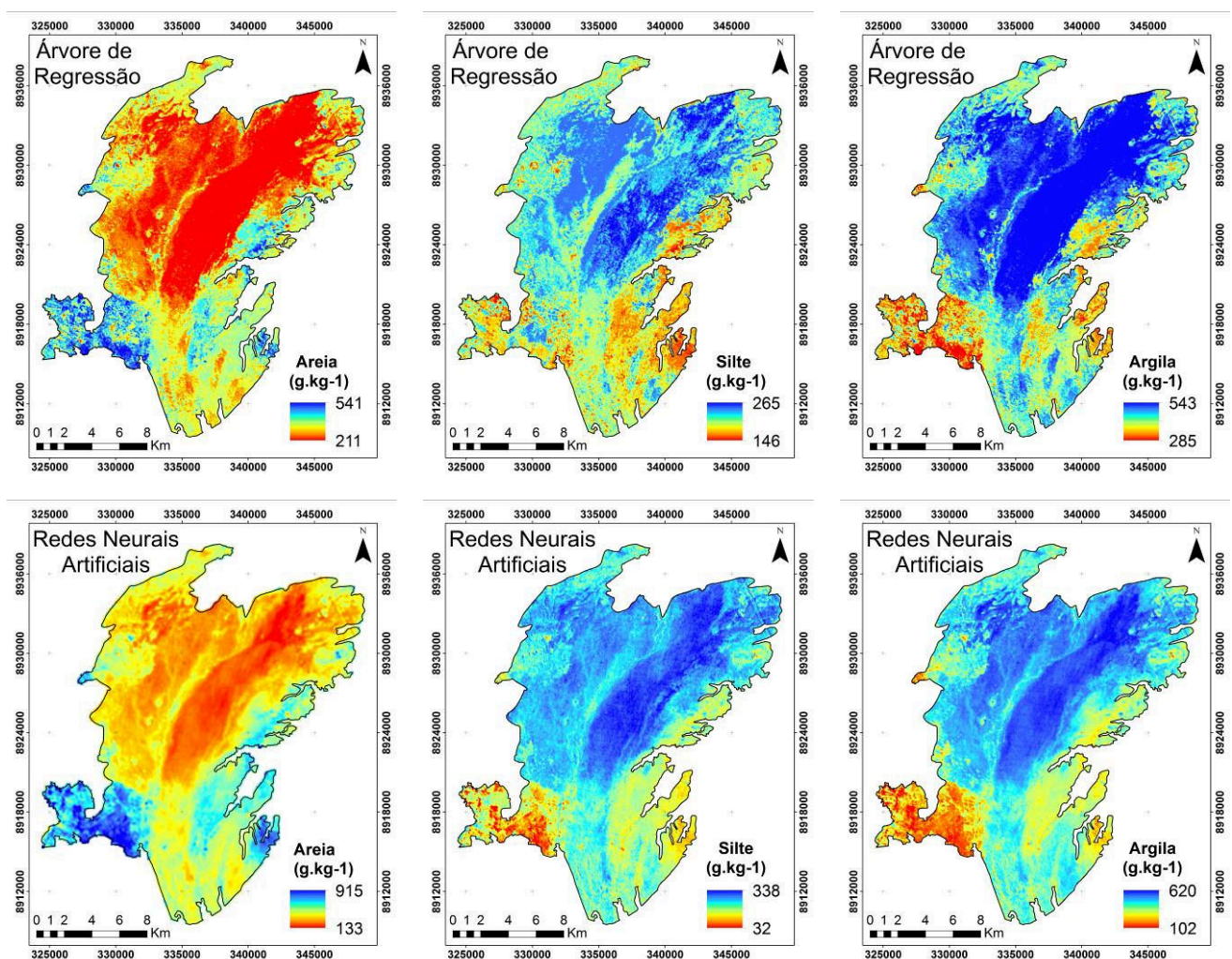


Figura 1. Distribuição espacial dos atributos físicos estimados pelos métodos utilizados.