

# ANAIS

## XX RBMCSA REUNIÃO BRASILEIRA DE MANEJO E CONSERVAÇÃO DO SOLO E DA ÁGUA

O SOLO SOB AMEAÇA: CONEXÕES  
NECESSÁRIAS AO MANEJO E  
CONSERVAÇÃO DO SOLO E ÁGUA

20 as 24 de novembro de 2016

Foz do Iguaçu - PR

Editores

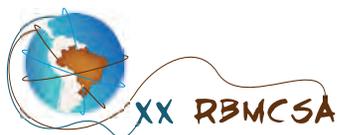
Arnaldo Colozzi Filho

João Henrique Caviglione

Graziela Moraes de Cesare Barbosa

Luciano Grillo Gil

Tiago Santos Telles



**Sociedade Brasileira de  
Ciência do Solo**  
**Núcleo Estadual Paraná**



NEPAR  
Curitiba  
2016

## MAPEAMENTO DIGITAL DE ATRIBUTOS FÍSICOS E FÍSICO-HÍDRICOS DO SOLO POR TÉCNICAS DE MINERAÇÃO DE DADOS

Cesar da Silva Chagas, Silvio Barge Bhering, Waldir de Carvalho Junior, Nilson Rendeiro Pereira

Embrapa Solos, Pesquisador, Rio de Janeiro - RJ, [chagas.rj@gmail.com](mailto:chagas.rj@gmail.com).

**Palavras-chave:** árvores de regressão; random forest; redes neurais artificiais.

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. Vários estudos têm demonstrado as relações entre diferentes atributos do solo e dados de sensores remotos, 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 redes neurais artificiais (RNA), random forest (RF) e árvores de regressão (AR) tem sido menos comum. O presente estudo tem por objetivo comparar a eficiência dos modelos AR, RF e RNA na predição de atributos físicos e físico-hídricos do solo utilizando dados do sensor TM do Landsat 5 como covariáveis ambientais em uma área do semiárido brasileiro.

O estudo foi realizado em uma área pertencente ao projeto de irrigação Salitre, no município de Juazeiro, Estado da Bahia. A área possui aproximadamente 35.000 ha e está localizada entre as coordenadas UTM 8.909.500 e 8.938.300 m N e 324.300 e 348.800 m E, zona 24S. O clima é semiárido, segundo Köppen, com vegetação de caatinga hiperxerófila, relevo plano e a geologia formada por calcários e rochas gnáissicas-graníticas. As classes de solo mais representativas são os Vertissolos, Cambissolos e Planossolos.

Para predição dos atributos areia, silte, argila, capacidade de campo (CC) e ponto de murcha permanente (PMP) foram utilizados dados da camada superficial (0 - 20 cm) de 399 perfis de solos disponibilizados pela CODEVASF. Foram utilizados como covariáveis dados do sensor TM do Landsat 5 (números digitais), órbita/ponto 217/67, com resolução espacial de 30 m, de janeiro de 2007, conforme se segue: b1 (0,450-0,515  $\mu\text{m}$ ), b2 (0,525-0,605  $\mu\text{m}$ ), b3 (0,630-0,690  $\mu\text{m}$ ), b4 (0,755-0,900  $\mu\text{m}$ ), b5 (1,550-1,750  $\mu\text{m}$ ), b7 (2,090-2,350  $\mu\text{m}$ ), índices NDVI ( $(b4 - b3)/(b4 + b3)$ ) e GSI ( $(b3 - b1)/(b3 + b2 + b1)$ ), e as relações  $b3/b2$ ,  $b3/b7$  e  $b5/b7$ . Os modelos de predição foram implementados no *software* R (R DEVELOPMENT CORE TEAM, 2013), através da função `rpart` (AR) e dos pacotes `randomForest` (RF) e `RSNNS` (RNA).

A importância das covariáveis ambientais para cada um dos atributos avaliados foi estimada através da correlação linear de Pearson, implementada no R (R DEVELOPMENT CORE TEAM, 2013), através da função `cor.test`, com valores inferiores a 0,05 indicando que a correlação é significativa. Em seguida, a performance dos modelos foi avaliada utilizando-se um conjunto de dados de validação independente (80 amostras), obtidos aleatoriamente, pelo cálculo da correlação entre os valores observados e os valores estimados através do coeficiente de determinação ( $R^2$ ).

Os resultados da análise da correlação linear de Pearson mostraram que, no geral, as covariáveis ambientais utilizadas apresentaram correlação significativa com os atributos dos solos estudados

( $p < 0,05$ ). A areia apresentou correlação significativa com a maioria das covariáveis utilizadas, com exceção das covariáveis b1 e b5/b7, enquanto o silte não apresentou correlação significativa com as covariáveis b1, b4 e b5/b7. Já o comportamento da argila com relação as covariáveis ambientais foi antagônico, porém de mesma magnitude, ao verificado para a areia, apenas as covariáveis b1 e b5/b7 não apresentaram correlação significativa com este atributo e as covariáveis mais relevantes foram b3/b7 ( $r = 0,56$ ), GSI ( $r = 0,40$ ), NDVI ( $r = -0,39$ ) e b3/b2 ( $r = 0,25$ ). A CC e o PMP tiveram o maior número de covariáveis não correlacionadas significativamente (b1, b2, b3 e b5/b7). Dessa maneira, apenas as covariáveis que tiveram correlação significativa com os atributos foram utilizadas nos modelos de predição.

Os resultados para o coeficiente de determinação ( $R^2$ ) obtidos pelos modelos de AR podem ser considerados como moderados para a areia (0,56), argila (0,52) e PMP (0,41); fraco para a CC (0,34) e muito fraco para o silte (0,17). 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 ( $R^2 = 0,44$ ).

Os modelos de RF tiveram desempenho para o  $R^2$  moderado para a areia (0,64), argila (0,56), CC (0,52) e PMP (0,42) e fraco para o silte (0,27). Akpa et al. (2014) relataram valores de  $R^2$ , na camada superficial do solo (0 - 15 cm), de 0,48 - 0,49 para areia, 0,26 - 0,27 para o silte e 0,53 - 0,56 para argila, em um estudo realizado na Nigéria. Por sua vez, os resultados obtidos pelos modelos de RNA foram moderados para areia (0,65), argila (0,67), CC (0,58) e PMP (0,53), e fraco para o silte (0,31). Os modelos de RNA foram superiores aos modelos de RF e AR na predição de todos os atributos avaliados, sendo que as maiores diferenças foram verificadas entre a RNA e a AR para a CC (0,24), argila (0,15) e silte (0,14). Viscarra Rossel e Behrens (2010), utilizaram métodos de mineração de dados, entre eles RNA, RF e AR 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  $R^2$  para RNA superiores aos demais modelos e aos encontrados no presente estudo.

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.

O modelo RNA apresentou coeficientes de determinação ( $R^2$ ), considerando a amostras de validação, superiores aos obtidos pelos modelos RF e 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.

## Referências

- BOETTINGER, J.L.; RAMSEY, R.D.; BODILY, J.M.; COLE, N.J.; KIENAST-BROWN, S.; NIELD, S.J.; SAUNDERS, A.M.; STUM, A.K. Landsat Spectral Data for Digital Soil Mapping. In: HARTEMINK, A. E.; MCBRATNEY, A. B.; MENDONÇA-SANTOS, M. L. (Eds.). **Digital Soil Mapping with limited data**. New York, Springer-Verlag, 2008. p.192-202.
- DEMATTÊ, J.A.M.; GALDOS, M.V.; GUIMARÃES, R.V.; GENÚ, A.M.; NANNI, M.R.; ZULLO JR, J. Quantification of tropical soil attributes from ETM+/LANDSAT-7 data. **Int. J. Remote Sens.**, v.28, p.3813-3829, 2007.

HENDERSON, B.L.; BUI, E.N.; MORAN, C.J.; SIMON, D.A.P. Australia-wide predictions of soil properties using decision trees. **Geoderma**, v.124, p.383-398, 2005.

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](http://www.r-project.org/isbn%203-900051-07-0)> . Acesso em 08 de maio de 2013.

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