



## Mapeamento Digital da CTC em Solos do Semiárido Brasileiro

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:** O planejamento de uso e o manejo sustentável da terra requerem informações confiáveis sobre a distribuição espacial de propriedades físicas e químicas do solo que afetam os processos e serviços na paisagem. Neste sentido, a capacidade de troca catiônica (CTC) é uma das propriedades do solo mais importantes, sendo considerada como um indicador vital da qualidade do solo, porém, sua obtenção é cara e demorada. Embora muitos estudos tenham sido conduzidos para identificar o padrão espacial da distribuição de propriedades do solo em várias escalas e várias paisagens, pouco se conhece sobre as relações existentes entre estas propriedades e diferentes covariáveis ambientais no semiárido brasileiro. Sendo assim, o objetivo desse estudo foi avaliar a eficiência dos modelos random forest (RF) e a cokrigagem (CK), implementados no software livre R, em conjunto com dados do sensor TM do Landsat 5 e atributos do terreno, no mapeamento digital da CTC em solos no semiárido brasileiro. Foram utilizadas 499 amostras da camada superficial (0 - 20 cm) dos solos. Os resultados obtidos, com base no conjunto de validação, mostraram que a melhor predição da CTC foi obtida pela co-krigagem, com  $R^2$  de 0,58 contra um  $R^2$  de 0,44 obtido pelo modelo random forest. O uso de covariáveis obtidas por sensoriamento remoto orbital e atributos do terreno, aliadas com as abordagens empregadas mostrou bom potencial para estimar valores de CTC nas condições do semiárido brasileiro.

**Termos de indexação:** mapeamento digital de solos, random forest, cokrigagem.

### INTRODUÇÃO

A implementação de uma agricultura sustentável, bem como o manejo ambiental requer um melhor entendimento sobre a variabilidade dos atributos solo, em escalas adequadas. Neste sentido, a capacidade de troca catiônica (CTC), importante propriedade do solo e considerada como um indicador vital de sua qualidade, é de obtenção cara e demorada.

Dados de sensores remotos orbitais podem ser utilizados como covariáveis ambientais no

mapeamento digital dos solos, especialmente em regiões áridas e semiáridas (Boettinger et al., 2008), facilitando, dessa maneira, o mapeamento e reduzindo a necessidade de levantamentos de campo onerosos e demorados.

Segundo McBratney et al. (2003), imagens de sensores remotos são a principal fonte de dados para o mapeamento digital de solos e tem sido usadas, principalmente para representar as variáveis da paisagem que estão relacionadas com os fatores de formação dos solos, como a vegetação e material de origem (covariáveis do solo).

Vários estudos têm demonstrado as relações existentes entre diferentes atributos do solo e dados de sensores remotos (Demattê et al., 2007; Liao et al., 2013). Entre os métodos mais comumente utilizados para a predição desses atributos destacam-se o geoestatísticos como a krigagem e a cokrigagem (Bishop & McBratney, 2001; Liao et al., 2011). A aplicação de métodos de mineração de dados como RF (Lagacherie et al., 2013; Vaysse & Lagacherie, 2015), tem sido menos comum.

O presente estudo tem por objetivo comparar a eficiência dos modelos RF e CK na predição da CTC da camada superficial dos solos, utilizando dados do sensor TM do Landsat 5 e atributos do terreno 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 uma área de aproximadamente 35000 ha, pertencente ao projeto de irrigação Salitre, localizada entre as coordenadas UTM 8.909.500 e 8.938.300m N e 324.300 e 348.800m E, zona 24S, no município de Juazeiro, Estado da Bahia. O clima na região é do tipo BSwh' (clima semiárido com inverno seco e verão chuvoso), segundo Köppen e 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 CTC, determinada conforme Embrapa (1979), foram utilizados dados da camada superficial (0-20 cm) de 499 perfis de solos disponibilizados pela CODEVASF, coletados no ano de 1989. Como covariáveis foram utilizados 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), e os atributos do terreno elevação (Elev), índice topográfico de umidade (TWI), declividade relativa (RSP), distância vertical do canal (VDCN) e profundidade do vale (VD).

### Modelos de predição

Neste estudo foram utilizados como modelos preditivos *random forest* (RF) e a cokrigagem (CK), implementados no R (R Development Core Team, 2013), com a utilização dos pacotes randomForest (RF) e gstat (CK). RF é uma técnica não paramétrica que consiste de uma combinação de muitas árvores preditoras (floresta), na qual cada árvore é gerada a partir de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores na floresta (Breiman, 2001). As subdivisões dentro de cada árvore são determinadas com base em um subconjunto de variáveis preditoras escolhido aleatoriamente a partir do total de preditores existentes. Os parâmetros de configuração utilizados foram: número de árvores na floresta (*ntree*) igual a 500 e número de variáveis em cada árvore (*mtry*) igual a 4 (Liaw & Wiener, 2002).

A cokrigagem, uma importante técnica geoestatística, é uma extensão da krigagem na qual uma variável primária é predita com base na correlação com uma ou mais variáveis secundárias e na co-dependência espacial existente (McBratney & Webster, 1983). A performance dos modelos foi medida, usando amostras de validação independentes (124 amostras que correspondem a aproximadamente 25% do conjunto total), e que não foram utilizadas na fase de treinamento (375 amostras), pelo cálculo da correlação entre os valores observados e os valores estimados através do coeficiente de determinação ( $R^2$ ).

## RESULTADOS E DISCUSSÃO

Uma das vantagens da RF é que ela pode ser utilizada adicionalmente como um método de análise para estimar a importância das variáveis ambientais, conforme mostrado na **figura 1**. Nesta análise as covariáveis mais importantes foram, em ordem decrescente, a relação B3/B7, índice NDVI, B1, relação B5/B7, B4, B7, B5, B3/B2, B2, Elev e B3, todas com importância superior a 10%. Oliveira et al. (2012) utilizaram este tipo de análise para selecionar as covariáveis mais importantes para compor um modelo de predição da ocorrência de fogo na Europa Mediterrânea usando como critério o limiar de 20%.

Por outro lado, os atributos do terreno RSP, Valley, VDCN e TWI tiveram importância bastante reduzida e foram eliminados do modelo final. A importância reduzida destas covariáveis pode ser explicada pelo fato da área apresentar relevo predominantemente plano, no qual a habilidade dos atributos do terreno na predição de propriedades do solos é limitada.

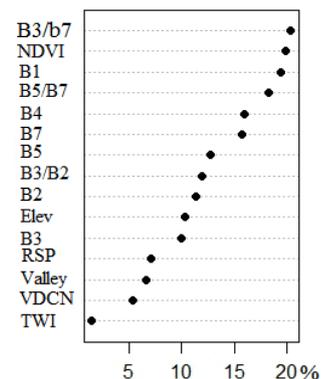


Figura 1. Importância das covariáveis ambientais estimada pelo modelo RF para CTC.

O resultado obtido pelo modelo RF com a utilização das covariáveis selecionadas, considerando o conjunto de amostras de validação independente foi de 0,44 para o coeficiente de determinação ( $R^2$ ). Levando-se em conta as características dos dados utilizados estes resultados podem ser considerados como moderadamente satisfatórios.

Lagacherie et al. (2013) utilizaram RF, atributos do terreno derivados de um mde com resolução de 30 m e dados hiperespectrais no visível e infravermelho próximo na predição da CTC e obtiveram resultados de  $R^2$ , para as camadas estudadas, que variaram de 0,35 (60-100cm) e 0,79 (15-30cm), desta maneira, os resultados obtidos no presente estudo podem ser considerados como similares.

Por outro lado, foram superiores aos alcançados por Vaysse & Lagacherie (2015), que não obtiveram resultados satisfatórios na predição da CTC usando RF. Neste caso, a baixa performance foi atribuída a pequena variação de escala determinada pelo material de origem e a relação erosão/deposição ao longo da pendente que não pode ser capturada pela resolução espacial das covariáveis utilizadas (100m). Além disso, os autores ressaltaram a necessidade de melhora na qualidade do conjunto de dados de solos, visando a melhoria do desempenho da predição.

Na **figura 2** é apresentada a distribuição espacial da CTC obtida com a aplicação do modelo RF. Os teores de CTC variaram de 5,7 a 61,4  $\text{cmol}_c.\text{kg}^{-1}$ , nas amostras de treinamento, desta maneira, inerente ao modelo, o RF produziu valores de distribuição espacial desse atributo dentro da amplitude do valor original.

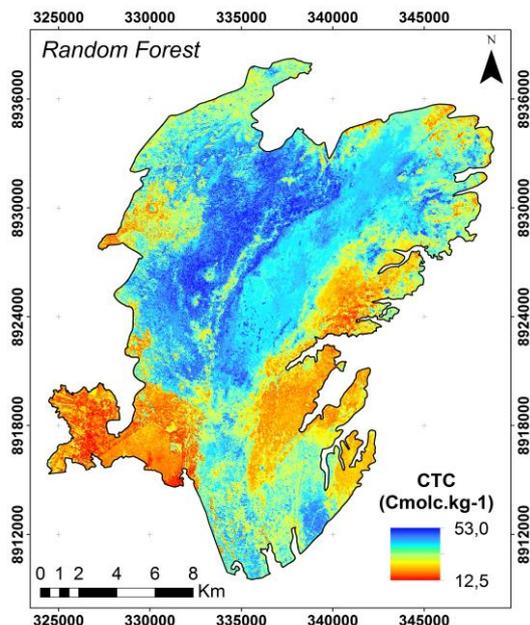


Figura 2. Distribuição espacial da CTC estimada pelo random forest.

As covariáveis mais importantes selecionadas pelo modelo RF foram também utilizadas na cokrigagem. O semivariograma obtido para a CTC do solo fornece uma descrição da sua dependência espacial e alguns indicativos sobre os possíveis processos que afetam sua distribuição (Liao et al., 2011).

Conforme mostrado na **figura 3** o modelo exponencial ajustou de modo satisfatório o semivariograma, que apresenta efeito pepita de 12,77, patamar igual a 54,43, relação pepita/patamar de 23,5% e alcance de 536 m. O

valor do efeito pepita maior do que zero pode ser causado pelo erro de medição ou pela variabilidade espacial da propriedade não explicada (Liao et al, 2013). A relação pepita/patamar (23,5%) inferior a 25%, evidencia a forte dependência espacial da CTC na área e pode estar indicando que a variabilidade espacial desse atributo é controlada principalmente por fatores intrínsecos como o material de origem dos solos (Cambardella et al., 1994).

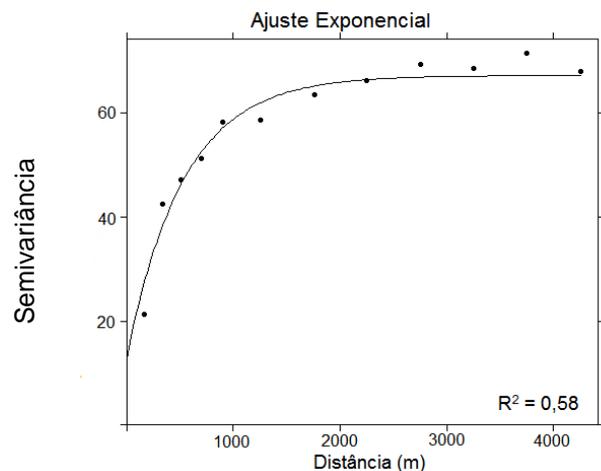


Figura 3. Semivariograma da CTC obtido pela cokrigagem.

Em seguida, este semivariograma foi utilizado para interpolar os valores da CTC espacialmente e o coeficiente de determinação ( $R^2$ ) de 0,58, obtido considerando-se o conjunto de amostras de validação independente, foi superior ao obtido pela RF ( $R^2 = 0,44$ ). O resultado obtido com a cokrigagem foi superior ao encontrado por Liao et al. (2013), que utilizaram a cokrigagem e componentes principais na predição da CTC em solos da China.

Na **figura 4** nota-se que a predição pela CK produziu um mapa com um intervalo de valores maior do que aquele produzido pela RF (**Figura 2**) e abrangendo uma amplitude mais próxima do valor real das amostras utilizadas no treinamento, variando de 9,8 a 65,3  $\text{cmol}_c.\text{kg}^{-1}$ , e indicando, segundo Liao et al. (2011), que a CK produziu predições mais acuradas do que a RF.

Os maiores teores de CTC estão concentrados na parte central da área de estudo onde são encontrados os Vertissolos, enquanto os teores mais baixos estão relacionados com Cambissolos de textura média que foram identificados na parte sudoeste da área.

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.

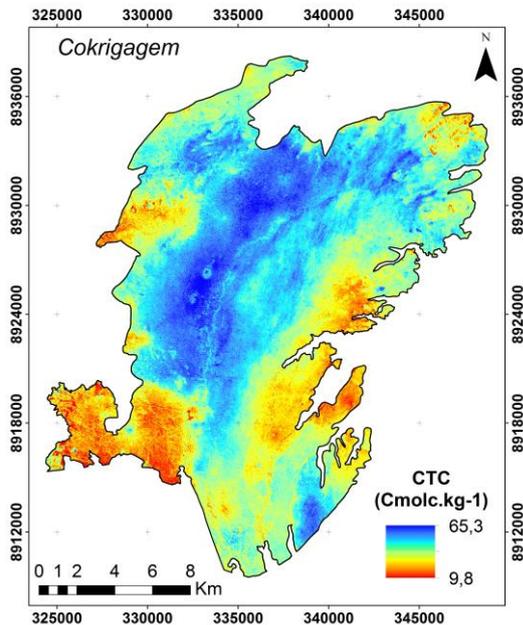


Figura 4. Distribuição espacial da CTC estimada pela cokrigagem.

## CONCLUSÕES

A cokrigagem apresentou coeficientes de determinação ( $R^2$ ), considerando as amostras de validação, superiores aos obtidos pelo modelo RF na predição da CTC.

As covariáveis mais importantes foram: a relação B3/B7, índice NDVI, B1, relação B5/B7, B4, B7, B5, B3/B2, B2, Elev e B3, que foram utilizadas tanto no modelo final da RF quanto da CK.

Os resultados obtidos para o semivariograma mostraram que a CTC apresenta uma forte dependência espacial na área estudada.

Embora os resultados obtidos pela cokrigagem sejam satisfatórios acredita-se que a melhoria da resolução espacial das covariáveis possam melhorar ainda mais o desempenho da predição da CTC na área.

## REFERÊNCIAS

BISHOP, T.F.A. & McBRATNEY, A.B. A comparison of prediction methods for the creation of field-extent soil property maps. *Geoderma*, 103:149-160, 2001.

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.

BREIMAN, L. Technical Report for V.3. Disponível em: <http://oz.berkeley.edu/users/breiman/randomforest2001.pdf>. Acesso 28 dez. 2014.

CAMBARDELLA, C. A. et al. Field-scale variability of soil properties in central Iowa soils. *Soil Sci. Soc. Am. J.*, v.58, p.1501-1511, 1994.

DEMATTE, 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.

LAGACHERIE, P. et al. Combining Vis-NIR hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia). *Geoderma*, v. 209, p.168-176, 2013.

LIAO, K. et al. Cokriging of soil cation exchange capacity using the first principal component derived from soil physico-chemical properties. *Agricultural Sciences in China*, v. 10, p.1246-1253, 2011.

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.

LIAN, A. & WIENER, M. Classification and regression by randomForest. *R News* 2(3):18-22, 2002.

McBRATNEY, A. B. & WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties. *Journal of Soil Science*, v. 34, p. 137-162, 1983.

McBRATNEY, A. B. et al. On digital soil mapping. *Geoderma*, v. 117, p. 3-52, 2003.

OLIVEIRA, S. et al. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. *Forest Ecology and Management*, v. 275, p. 117-129, 2012.

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).

VAYSSE, K. & LAGACHERIE, P. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional*, v.4, p.20-30, 2015.