



## Influência do modelo digital de elevação no mapeamento digital de atributos do solo por random forest (RF)

**Silvio Barge Bhering<sup>(1)</sup>; César da Silva Chagas<sup>(1)</sup>; Waldir de Carvalho Junior<sup>(1)</sup>; Nilson Rendeiro Pereira<sup>(1)</sup>; Braz Calderano Filho<sup>(1)</sup>**

<sup>(1)</sup> Pesquisador; Embrapa Solos; Rio de Janeiro; [silvio.bhering@embrapa.br](mailto:silvio.bhering@embrapa.br); [cesar.chagas@embrapa.br](mailto:cesar.chagas@embrapa.br); [waldir.carvalho@embrapa.br](mailto:waldir.carvalho@embrapa.br); [nilson.rendeiro@embrapa.br](mailto:nilson.rendeiro@embrapa.br); [braz.calderano@embrapa.br](mailto:braz.calderano@embrapa.br)

**RESUMO:** As propriedades físicas e químicas dos solos são importantes para o manejo dos solos, pois influenciam a sua fertilidade, as taxas de infiltração, retenção de umidade, entre outros. Muito da variabilidade dessas propriedades pode ser explicada por atributos do terreno derivados de modelos digitais de elevação (MDE). Todavia, as cartas topográficas, utilizadas para geração destes modelos, ainda são uma grande carência no Brasil, seja em cobertura seja em escala. Dessa forma, a utilização de dados do “Shuttle Radar Topography Mission” - SRTM pode contribuir no mapeamento dessas propriedades do solo. O objetivo desse estudo foi avaliar a influência da fonte do modelo digital de elevação na predição por *random forest* (RF) dos teores de areia, argila e capacidade de troca catiônica (CTC). Nessa avaliação foram utilizadas amostras superficiais (0-20cm) de 44 perfis de solos, coletados durante a elaboração do Zoneamento agroecológico do Estado do Mato Grosso do Sul, e como fonte de dados para elaboração dos MDEs, cartas topográficas do IBGE e dados do SRTM (Topodata). Complementarmente, utilizou-se também dados do sensor TM do Landsat 5. Os valores do coeficiente de determinação ( $R^2$ ) obtidos com essa análise foram superiores, para todos os atributos avaliados, quando se utilizou os dados do SRTM (Topodata). A moderada performance do modelo *random forest* (RF) neste estudo pode ser atribuída ao tamanho reduzido (44 perfis) do conjunto de dados utilizado.

**Termos de indexação:** pedometria, morfometria, rio Formoso

### INTRODUÇÃO

A caracterização e distribuição dos atributos do solo são de suma importância no planejamento de uso e ocupação das terras. Por outro lado, a carência dessas informações associado ao uso inadequado dos solos pode induzir a degradação dos ecossistemas e elevar os níveis de emissão de carbono para a atmosfera (Mulder et al., 2011).

Propriedades físicas e químicas dos solos são muito importantes para o manejo dos solos, influenciando a sua fertilidade, as taxas de infiltração e a retenção de umidade. Entre as variáveis

ambientais que representam os fatores de formação do solo, o relevo tem sido muito utilizado. Muito da variabilidade dessas propriedades pode ser explicada por atributos do terreno derivados dos modelos digitais de elevação (MDE).

A fonte de dados mais comum para a geração de MDEs no Brasil ainda são as curvas de nível e, de forma complementar, a rede hidrográfica, obtidas em cartas topográficas, principalmente aquelas elaboradas pelo Instituto Brasileiro de Geografia e Estatística – IBGE; entretanto, as escalas normalmente disponíveis não são adequadas para estudos mais detalhados; por esta razão, dados oriundos de imagens de sensores remotos estão sendo cada vez mais utilizados para a geração desses modelos (Chagas et al., 2010).

Todavia, cartas topográficas em escalas compatíveis é uma grande carência no Brasil. Dessa forma, a utilização de dados do “Shuttle Radar Topography Mission” - SRTM pode contribuir para o mapeamento dessas propriedades do solo.

O objetivo desse estudo foi avaliar a influência da fonte do modelo digital de elevação (MDE) na predição por *random forest* (RF) dos teores de areia, argila e capacidade de troca catiônica no horizonte A dos solos na bacia do rio Formoso, Bonito (MS), utilizando uma base de dados de solos limitada.

### MATERIAL E MÉTODOS

O estudo foi realizado na bacia do Rio Formoso, Município de Bonito (MS) com área de 1263 Km<sup>2</sup>. Segundo a classificação climática de Köppen, o clima nessa região é Aw (clima tropical seco e megatérmico com estação seca definida), temperatura média de 23,1°C e precipitação anual média próxima a 1.500 mm. A vegetação original da bacia era composta por floresta tropical subcaducifólia, floresta de várzea e por áreas com cerrado e área de transição. Atualmente, verifica-se que apenas uma pequena porcentagem da bacia ainda apresenta vegetação natural. A litologia predominantemente na área é formada por calcáreos, mármores e rochas pelíticas.

### Fonte de dados para elaboração do MDE

Os MDE's podem ser obtidos pela interpolação de feições vetoriais (pontos cotados, curvas de nível e



hidrografia) ou diretamente de sensores remotos. Nesse estudo, utilizaram-se como fonte de dados para a obtenção dos MDEs pontos cotados, curvas de nível equidistantes de 40 metros e hidrografia, contidos nas cartas topográficas do IBGE, na escala 1:100.000 (MDE Carta) e dados do Topodata (Valeriano, 2005) derivados do “Shuttle Radar Topography Mission” - SRTM, com resolução espacial de 30 metros (MDE Srtm). O MDE Carta foi gerado com a ferramenta *Topo to Raster* do ArcGIS Desktop v.10, que implementa um método de interpolação específico e hidrologicamente consistente.

#### Atributos do solo e covariáveis ambientais

Na análise e predição da areia, argila e CTC dos solos foram utilizados os dados da camada superficial (0 - 20 cm) de 44 perfis de solos coletados durante o Zoneamento Agroecológico do Estado do Mato Grosso do Sul. A composição granulométrica foi determinada pelo método do densímetro e separação das frações em areia (2-0,05 mm); e argila (<0,002 mm), e a CTC pela soma dos teores de Ca<sup>++</sup>, Mg<sup>++</sup>, Na<sup>+</sup>, K<sup>+</sup>, H+Al, conforme Embrapa (1997).

Foram utilizados como covariáveis ambientais atributos morfométricos derivados dos MDEs testados (MDE Carta e MDE Srtm) e dados do sensor TM do Landsat 5 (números digitais) com resolução espacial de 30m. Os atributos morfométricos foram: altitude, declividade, aspecto, curvatura, plano e perfil de curvatura, nível base do canal de drenagem (CNBL), índice de convergência (CI), fator LS, posição relativa da declividade (RSP), índice topográfico de umidade (TWI), profundidade do vale (VD) e distância vertical da rede de drenagem (VDCN), obtidos através de processamento no ArcGIS Desktop v. 10.1 e no Saga Gis. Os dados do Landsat 5 foram: 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  $\mu\text{m}$ ), índice NDVI ( $b_4 - b_3 / b_4 + b_3$ ) e as relações entre bandas B3/B2, B3/B7 e B5/B7, conforme Malone et al. (2009).

A importância das covariáveis ambientais foi estimada através da análise da correlação linear de Pearson e somente as covariáveis que apresentaram correlação significativa ( $p < 0,05$ ) com os atributos avaliados foram utilizadas.

#### Modelo de predição

Nesse estudo foi utilizada a *random forest* (RF) que é uma técnica não paramétrica desenvolvida por Breiman (2001) como uma extensão do programa CART (*Classification and Regression Trees*), para melhorar o desempenho de predição do modelo, que consiste de uma combinação de muitas

árvores predictoras (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. No caso da aplicação da RF para regressão o resultado final consiste da média dos resultados de todas as árvores (Breiman, 2001).

As RFs foram implementadas utilizando-se o pacote *randomForest* do R (R Development Core Team, 2013). O padrão para  $n_{tree}$  definido no sistema é de 500. Como valor de *nodesize* foi utilizado o padrão para os estudos de regressão que é de cinco para cada nó terminal. Com relação ao  $m_{try}$ , para problemas de regressão, o valor padrão estipulado é de um terço do número total de variáveis predictoras, deste modo, foi utilizado um  $m_{try}$  de três, para um número de variáveis predictoras igual a oito, conforme Liaw & Wiener (2002).

A performance dos modelos foi medida pelo cálculo da correlação entre os valores observados e os valores estimados através do coeficiente de determinação ( $R^2$ ).

## RESULTADOS E DISCUSSÃO

A estatística descritiva dos atributos da camada superficial (0 a 20 cm) dos solos é apresentada no **quadro 1**.

**Quadro 1** - Estatística descritiva dos atributos do solo.

Atributo	Max	Min	Méd	DP	CV (%)
Areia (g/kg)	862,0	58,0	412,7	206	50
Argila (g/kg)	724,0	40,0	325,3	177	54
CTC (cmol/kg <sup>-1</sup> )	40,2	1,5	12,5	9	74

Os valores de coeficiente de variação (CV) superiores a 50% em todos os casos, caracterizam a heterogeneidade do conjunto de amostras. A capacidade de troca catiônica foi o atributo que apresentou maior CV.

A análise da correlação de Pearson mostrou que as covariáveis correlacionadas com os atributos do solos avaliados, quando se utilizou o MDE Srtm, foram: Altitude, CNBL, RSP, TWI, B3, B5, B7, B3/B7, B5/B7 e NDVI para a areia; Altitude, CNBL, RSP, TWI, B1, B5 e B3/B7, para argila; e Altitude, declividade, CNBL, fator LS, RSP, VD, B2, B3, B4, B5, B7, B3/B2, B3/B7, B5/B7 e NDVI, para a CTC.

Para o MDE Carta as covariáveis correlacionadas foram as seguintes: areia – altitude, CNBL, B3, B5, B7, B2/B7, B5/B7, NDVI; argila - altitude, CNBL, TWI, B1, B5, B3/B7; e CTC- altitude, plano de



curvatura, CNBL, TWI, VD, B2, B3, B4, B5, B7, B3/B2, B3/B7, B5/B7, NVDI.

O **quadro 2** apresenta os resultados do coeficiente de determinação ( $R^2$ ) obtidos na predição dos atributos de acordo com a fonte de dados do modelo digital de elevação utilizado.

**Quadro 2-** Coeficiente de determinação para os atributos avaliados

Atributo	MDE CARTA	MDE SRTM	Diferença
	$R^2$		
Areia	0,22	0,33	0,11
Argila	0,31	0,49	0,18
CTC	0,35	0,40	0,05

Os valores do coeficiente de determinação ( $R^2$ ) diferem entre em função da fonte de dados utilizadas na geração do MDE, sendo superior, para todos os atributos, quando da utilização do MDE Srtm. O maior coeficiente de correlação ( $R^2$ ) foi obtido para o atributo argila, que inclusive mostra a maior diferença em relação ao resultado do MDE Carta (0,18) (Quadro 2).

Demattê et al. (2007) atribuíram os bons resultados de predição de areia e argila obtidos em seu estudo a alta variância e desvio padrão dos atributos dos solos e dos dados espectrais utilizados, em comparação com outra área de menor variância dos dados. No presente estudo os atributos do solo apresentaram variância elevada (CV > 50%).

Os teores das frações avaliadas variaram, nas amostras de solo de 58 a 862 g.kg<sup>-1</sup> (areia), 40 a 724 g.kg<sup>-1</sup> (argila) e de 1,5 a 40,2 cmoc.kg<sup>-1</sup>. Assim, os modelos utilizados produziram valores de distribuição espacial desses atributos muito parecidos entre si e próximos dos valores originais (**Figura 1**). Deste modo, os teores de areia variaram de 157 a 678 g.kg<sup>-1</sup>, no MDE Srtm e de 139 a 665 g.kg<sup>-1</sup>, no MDE Carta. Com relação a argila, os teores variaram entre 140 e 601 g.kg<sup>-1</sup> (MDE Srtm) e entre 147 e 612 g.kg<sup>-1</sup> (MDE Carta). Por sua vez, a CTC variou de 5 a 29 cmolc.kg<sup>-1</sup>, no MDE Srtm e no MDE Carta de 6 a 27 cmolc.kg<sup>-1</sup>.

## CONCLUSÕES

A moderada performance do modelo *random forest* (RF) neste estudo pode ser atribuída ao tamanho reduzido (44 perfis) do conjunto de dados utilizado.

O modelo RF apresentou coeficientes de determinação ( $R^2$ ), superiores para o MDE SRTM na predição da argila, da areia e da CTC (0,49; 0,33 e

040; respectivamente) Estes resultados são similares aos obtidos em outros estudos.

O uso de covariáveis ambientais (morfológicas) derivadas do modelo digital de elevação obtido por sensoriamento remoto orbital (SRTM) aliado a abordagem *random forest* mostrou bom potencial de utilização para estimar valores de granulometria, areia e argila; e capacidade de troca catiônica utilizando uma base de dados de solos limitada.

## REFERÊNCIAS

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

CARVALHO JUNIOR, W.; LAGACHERIE, P.; CHAGAS, C. S.; CALDERANO FILHO, B.; BHERING, S. B. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma*, v. 232, p. 479-486, 2014.

CHAGAS, C.S; FERNANDES FILHO, E.I.; CARVALHO JUNIOR, W.; SOUZA NETO, N.C. Avaliação de modelos digitais de elevação para aplicação em um mapeamento digital de solos. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.14, n.2, p.218–226, 2010.

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. *International Journal of Remote Sensing*, v.28, p.3813-3829, 2007.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solo. Manual e métodos de análise do solo. Rio de Janeiro, 1997. 2ªed.

LIAW, A; WIENER, M. Classification and regression by randomForest. *R News* 2(3):18–22, 2002.

MALONE, B. P.; McBRATNEY, A. B.; MINASNY, B.; LASLETT, G. M. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, v.154, p.138-152, 2009.

MULDER, V. L.; DE BRUIN, S.; SCHAEPMAN, M. E.; MAYR, T. R. 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.

VALERIANO, M. M. Modelo digital de variáveis morfológicas com dados SRTM para o território nacional: o projeto TOPODATA. In: XII Simpósio Brasileiro de Sensoriamento Remoto, 2005, Goiânia, GO. Anais do XII Simpósio Brasileiro de Sensoriamento Remoto, 2005. p. 1-8.

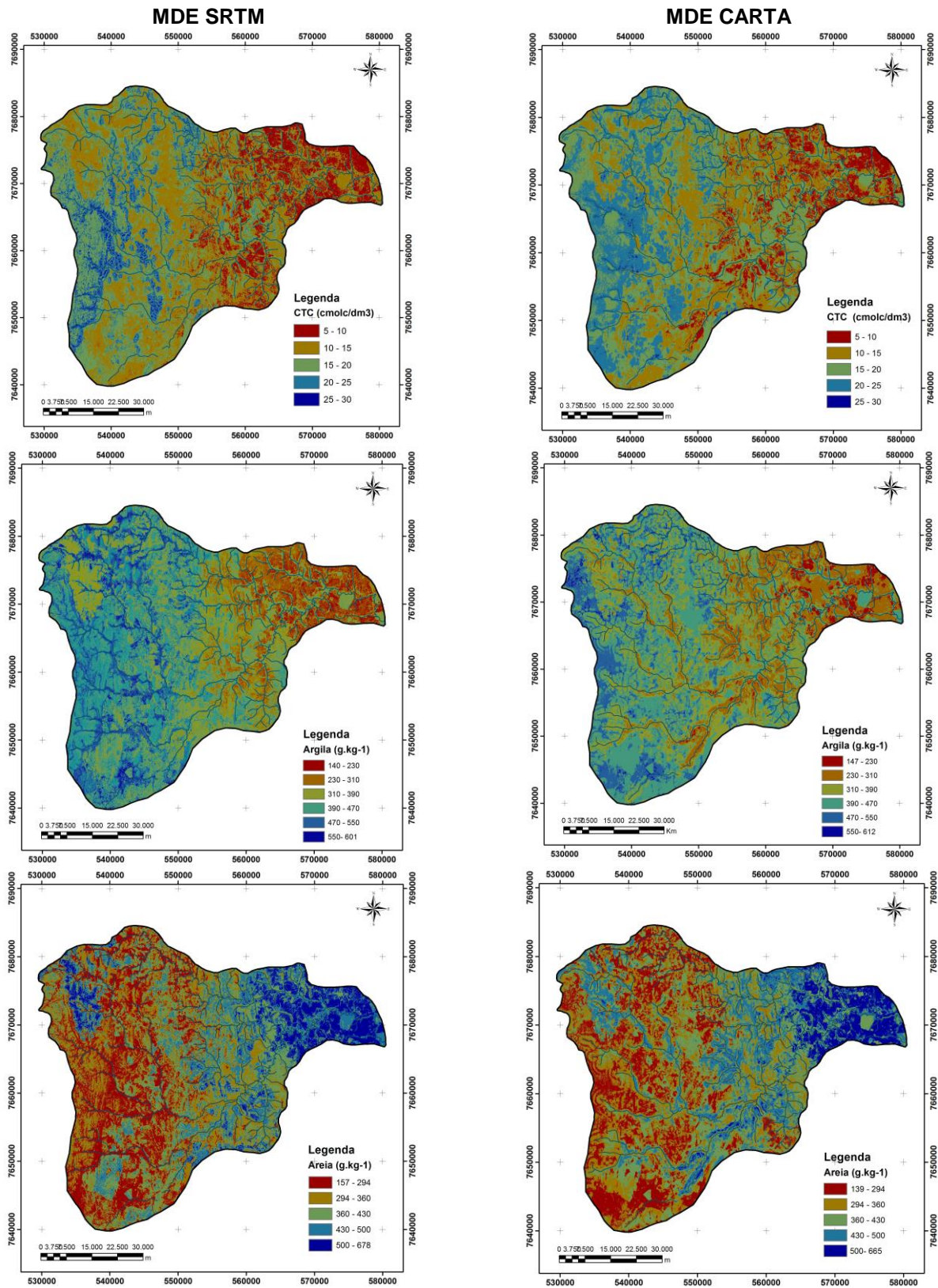


Figura 1 – Mapas da CTC, argila e areia gerados por random forest (RF) a partir de MDE Carta e MDE Srtm