



## Capacitação em Mapeamento Digital de Solos. Parte 2 – Estudo de Caso: Carbono do Solo em Campos dos Goytacazes, RJ<sup>(1)</sup>.

**Gustavo M. Vasques<sup>(2)</sup>; Ricardo O. Dart<sup>(3)</sup>; Jesus F. M. Baca<sup>(2)</sup>; Guillermo F. Olmedo<sup>(4)</sup>; Maria de Lourdes M. S. Brefin<sup>(2)</sup>.**

<sup>(1)</sup> Trabalho executado com recursos da Organização das Nações Unidas para a Alimentação e a Agricultura

<sup>(2)</sup> Pesquisador; Embrapa Solos; Rio de Janeiro, RJ; gustavo.vasques@embrapa.br; jesus.baca@embrapa.br; lourdes.mendonca@embrapa.br; <sup>(3)</sup> Analista; Embrapa Solos; Rio de Janeiro, RJ; ricardo.dart@embrapa.br; <sup>(4)</sup> Pesquisador; Instituto Nacional de Tecnología Agropecuaria; Mendoza, Argentina; folmedo@mendoza.inta.gov.ar.

**RESUMO:** No período de 24 a 28 de setembro de 2012 foi realizado na Embrapa Solos, Rio de Janeiro, RJ, o *Curso de Capacitación en Cartografía Digital de Suelos*, que contou com a presença de 17 pedólogos de 17 países da América Latina e Caribe. O objetivo do curso foi capacitar os pedólogos em mapeamento digital de solos utilizando os softwares livres SAGA, R e RStudio. No curso realizou-se, como estudo de caso, o mapeamento do carbono orgânico do solo na camada 0-10 cm no município de Campos dos Goytacazes, RJ. Apresentamos as etapas de treinamento realizadas durante o curso, incluindo a preparação de dados de solos e covariáveis ambientais, análise exploratória de dados, modelagem espacial, produção de mapas e validação dos resultados obtidos. A combinação dos softwares livres SAGA, R e RStudio mostrou-se bastante flexível para a realização do curso e mapeamento digital de carbono orgânico do solo, sendo recomendada para outros cursos e projetos de mapeamento digital de solos. A estrutura do curso mostrou-se adequada também para implementação através de ensino à distância (resultados apresentados em outro trabalho nestes anais). Para êxito de cursos e projetos similares, salientamos a importância da experiência prévia dos participantes com conceitos e técnicas de pedologia, estatística, geoprocessamento e linguagem de programação, bem como da existência de um banco de dados de solos organizado e disponível.

**Termos de indexação:** pedometria, regressão-kriging, ensino de pedologia.

### INTRODUÇÃO

Oportunidades de treinamento em mapeamento digital de solos (MDS) estão aquém da demanda atual por capacitação em técnicas que incorporam modelagem quantitativa e geotecnologia na ciência do solo tanto no Brasil como no restante da América Latina e Caribe (LAC), notadamente em pedologia e levantamentos de solo.

Ainda, a oferta de cursos que utilizam software livre na LAC, embora esteja aumentando, também não é suficiente para acompanhar a tendência

global de desenvolvimento e utilização de software livre em ciência, o que acaba gerando um entrave ao avanço da ciência do solo nesses países, especialmente em direção a abordagens mais quantitativas.

Nesse contexto, a Embrapa Solos realizou entre 24 e 28 de setembro de 2012 o *Curso de Capacitación en Cartografía Digital de Suelos*, com duração de 40 horas, em que participaram 17 pedólogos de 17 países da LAC. O curso teve conteúdo teórico-prático em espanhol, sendo que, para os exercícios práticos, foi realizado como estudo de caso o mapeamento digital do carbono orgânico do solo (Corg) no município de Campos dos Goytacazes, RJ.

Os softwares livres utilizados no curso foram o SAGA 2.0.8 (SAGA User Group Association, 2012) para geoprocessamento e o R 2.15.1 (R Core Team, 2012), implementado através do ambiente de desenvolvimento integrado RStudio 0.96.331 (RStudio, 2012), para análises estatísticas.

O objetivo desse trabalho é apresentar, de maneira sucinta, os principais módulos (etapas) do curso realizado e os produtos do mapeamento de Corg obtidos.

### MATERIAL E MÉTODOS

#### Área de estudo

O município de Campos dos Goytacazes (**Figura 1**), no norte do estado do Rio de Janeiro, foi utilizado como área de estudo do curso. Os critérios utilizados para escolha da área foram a existência de dados de solos e covariáveis ambientais, a familiaridade dos instrutores do curso com esses dados, e o tamanho relativamente reduzido da área (~4.049 km<sup>2</sup>).

#### Dados

Os dados de Corg utilizados no curso foram cedidos por um projeto anterior de mapeamento de Corg no estado do Rio de Janeiro (Mendonça-Santos et al., 2007). Foram selecionados 99 pontos amostrais contendo dados de Corg na camada 0-10 cm na área de estudo (**Figura 1**), dos quais 80

foram separados aleatoriamente para treinamento do modelo e geração de mapas, e 19 foram utilizados exclusivamente para validação dos mesmos (**Tabela 1**).

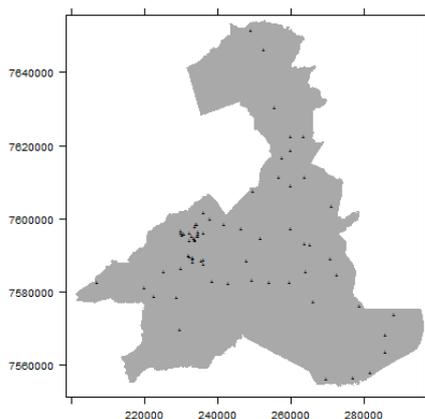
**Tabela 1** – Estatística descritiva do carbono orgânico do solo (g/kg) na camada 0-10 cm.

Conjunto	N <sup>1</sup>	Mínimo	Média	Máximo
Total	99	0,8	18,5	112,4
Treinamento	80	0,8	19,4	112,4
Validação	19	2,0	14,8	30,5

<sup>1</sup> Número de observações.

Dezenove variáveis independentes (covariáveis) foram preparadas no software SAGA e extraídas para os pontos amostrais de solo para serem utilizadas para predição de Corg. As 19 covariáveis incluíram seis bandas e o índice de vegetação por diferença normalizada (NDVI) do sensor *Thematic Mapper* do satélite Landsat e 12 variáveis de relevo obtidas a partir de um modelo digital de elevação (MDE) da *Shuttle Radar Topography Mission* (**Tabela 2**), todas na resolução espacial de 90 m.

Os dados foram processados na projeção *Universal Transverse Mercator* fuso 24 sul, datum *World Geodetic System 1984*.



**Figura 1** – Delineamento amostral em Campos do Goytacazes, RJ. Mapa produzido no R.

### Modelagem espacial

O mapa de Corg foi gerado, a partir das 80 observações de treinamento, através de regressão-krigagem, realizada em três etapas. Primeiro, na etapa de *modelagem da tendência global*, o Corg foi predito através de regressão linear múltipla (RLM) em função das 19 covariáveis ambientais, utilizando seleção de variáveis passo a passo (*stepwise*) com valor-p de 0,10.

Segundo, na etapa de *modelagem da tendência*

*local*, os resíduos da RLM foram interpolados através de krigagem ordinária pontual. Para isso, a existência de autocorrelação espacial nos resíduos foi confirmada visualmente nos semivariogramas.

Terceiro, para *produção do mapa final de Corg*, foram somados os mapas de Corg predito (etapa 1) e de resíduos da RLM interpolados (etapa 2). O modelo de RLM e o mapa final de Corg foram validados através do cálculo do coeficiente de determinação ( $R^2$ ) e do erro padrão dos resíduos ( $S_{res}$ ) utilizando as 19 observações de validação.

No curso, na etapa de *modelagem da tendência global*, foram também utilizados métodos de predição não-paramétricos (redes neurais artificiais e árvores de regressão), cujos resultados não são apresentados aqui.

**Tabela 2** – Covariáveis ambientais utilizadas para predição de carbono orgânico do solo.

Landsat	Relevo <sup>1</sup>
Banda 1,	DEM [no sinks], Analytical
Banda 2,	Hillshading, Slope, Aspect, Plan
Banda 3,	Curvature, Profile Curvature,
Banda 4,	Convergence Index, Catchment
Banda 5,	Area, Wetness Index, LS-Factor,
Banda 7,	Altitude above Channel Network,
NDVI	Channel Network Base Level

<sup>1</sup> Nomes originais gerados pelo software SAGA.

## RESULTADOS E DISCUSSÃO

O curso iniciou-se com a instalação dos softwares e organização da estrutura de dados do curso no computador. Somente foram fornecidos aos participantes os dados de solos previamente organizados, seis bandas Landsat e o MDE.

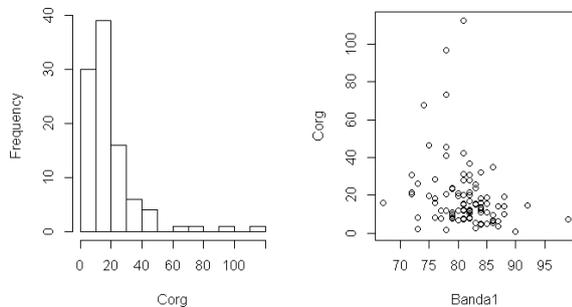
Em seguida, três módulos introdutórios foram apresentados, sendo: *Introdução ao MDS*; *Introdução ao R*; e *Introdução ao SAGA*. Então, seguiram-se os módulos de modelagem espacial e mapeamento (regressão-krigagem) propriamente ditos, cujos principais resultados são apresentados nos próximos itens.

### Preparação de dados

A preparação das covariáveis ambientais no software SAGA constituiu uma das etapas iniciais do curso. Foram utilizadas rotinas para: carregar e importar dados; definir projeção e reprojetar; remover depressões espúrias do MDE; gerar derivadas de relevo; reamostrar camadas Landsat de 30 para 90 m; calcular NDVI; extrair valores de pixels para pontos; e exportar dados.

### Análise exploratória de dados

Inicialmente, foram apresentados conceitos sobre organização e qualidade de dados. Então, métodos clássicos de análise exploratória e correção de dados (por exemplo, checagem de erros, estatística descritiva, etc.) e gráficos (**Figura 2**) foram implementados no R, utilizando, entre outras, as seguintes funções: *class*, *summary*, *plot*, *hist* e *boxplot*. Em seguida, foram apresentadas ferramentas interativas de análise exploratória de dados no R utilizando os pacotes *iPlots*, *JGR* e *Deducer* (que carrega o pacote *ggplot2* para geração de gráficos de alta qualidade).



**Figura 2** – Exemplos de gráficos no R: histograma (esquerda) e gráfico de dispersão (direita).

### Modelagem da tendência global

O modelo de RLM foi gerado com as funções *lm* (regressão linear) e *step* (seleção passo a passo) no R. Os pressupostos da RLM foram checados utilizando funções como *outlierTest*, *vif* e *ncvTest*, disponíveis nos pacotes *car* e *HH*, sendo que foram identificados e retirados nove *outliers*. O modelo final de predição de Corg foi significativo ao nível de 0,05 de significância e apresentou um  $R^2$  de 0,53 e  $s_{res}$  de 6,3, resultado que consideramos satisfatório, dados todos os pressupostos envolvidos em MDS.

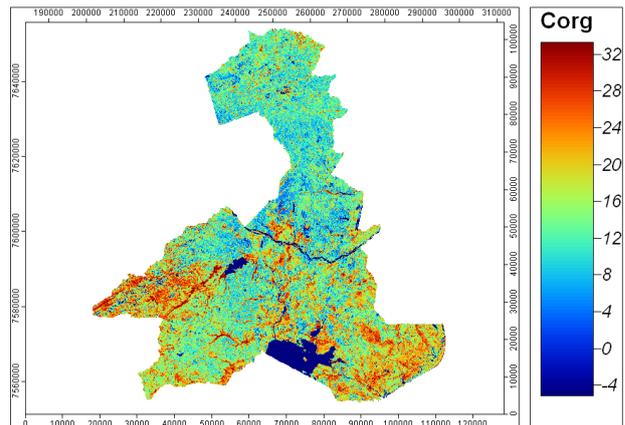
O mapa de tendência global do Corg (**Figura 3**) foi gerado com a ferramenta *Grid Calculator* no SAGA a partir da fórmula:  $Corg = -29,7 + 75,4 * Slope + 0,1 * Convergence.Index - 2E-07 * Catchment.Area + 1,9 * Wetness.Index - 3,7 * LS.Factor + 1,0 * Banda3 - 0,7 * Banda4 + 176,9 * NDVI$ . Ele refletiu a distribuição espacial das variáveis ambientais, notadamente NDVI e variáveis de relevo.

### Modelagem da tendência local

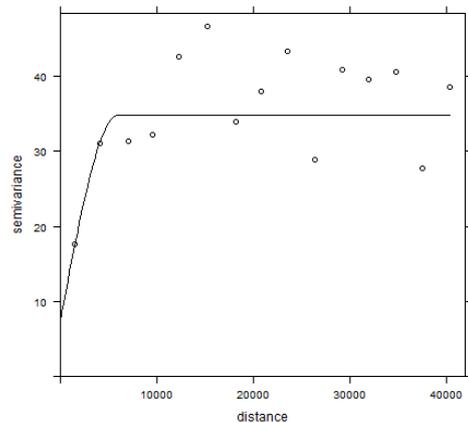
Nesse módulo foram apresentados alguns formatos de dados espaciais usados no R e funções para produção de mapas (**Figuras 1 e 5**). Foram utilizados os pacotes *lattice* e *gstat* (que carrega o pacote *sp* para trabalhar com dados espaciais).

Os resíduos da RLM do Corg ajustaram-se ao modelo de semivariograma esférico (efeito pepita:

7,2 (g/kg)<sup>2</sup>; patamar: 27,7 (g/kg)<sup>2</sup>; e alcance: 6.040 m) e apresentaram autocorrelação espacial positiva (**Figura 4**). A distribuição espacial dos resíduos da RLM (**Figura 5**) foi nitidamente influenciada pela distribuição dos pontos amostrais, característica inerente à krigagem.



**Figura 3** – Carbono orgânico do solo (g/kg) na camada 0-10 cm predito através de regressão linear múltipla (tendência global). Mapa produzido no SAGA.



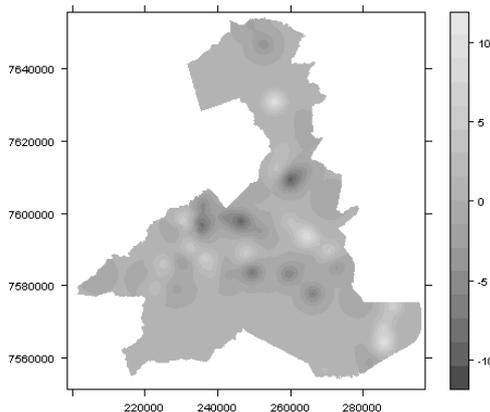
**Figura 4** – Semivariogramas empírico (pontos) e ajustado por modelo esférico (linha).

### Produção do mapa final de Corg

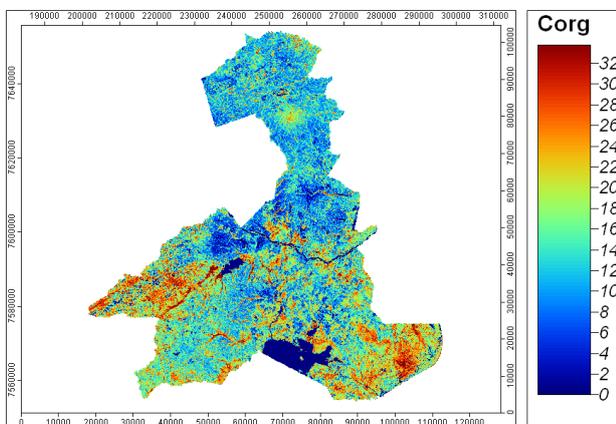
O mapa final de Corg (**Figura 6**) reproduz em maior grau a tendência global de distribuição espacial do Corg (**Figura 3**), pois esta contribui com maior magnitude de valores (-5,2 a 33,2 g/kg) do que a tendência local (-10,4 a 10,5 g/kg), gerada em função dos resíduos da RLM (**Figura 5**).

A amplitude de valores do Corg predito através de regressão-krigagem assemelhou-se àquelas dos conjuntos de calibração (0,80 a 45,6 g/kg, após eliminação de nove *outliers*) e validação (2,0 a 30,5 g/kg). A média de Corg em 0-10 cm (18,5 g/kg) foi maior do que as médias encontradas por Fontana et

al. (2001) para 0-20 cm também no município de Campos dos Goytacazes, RJ (5,4 a 13,4 g/kg). Essa maior concentração de Corg nos primeiros 10 cm do solo corrobora a tendência de diminuição de Corg em profundidade observada por Fontana et al. (2001) ao comparar o Corg entre 0-20 e 20-40 cm.



**Figura 5** – Resíduos da regressão linear múltipla de carbono orgânico do solo (g/kg) na camada 0-10 cm interpolados através de krigagem ordinária pontual (tendência local). Mapa produzido no R.



**Figura 6** – Mapa final de carbono orgânico do solo (g/kg) na camada 0-10 cm em Campos dos Goytacazes, RJ, obtido através de regressão-krigagem. Mapa produzido no SAGA.

### Validação do modelo e mapa final

O conceito de incerteza e diferentes índices de qualidade (ou incerteza) foram apresentados no curso. Considerando as observações de validação, o modelo de RLM e o mapa final de Corg obtiveram  $R^2$  próximo de 0,1 e  $s_{res}$  de 8,5 e 8,4, respectivamente. Além de evidenciar limitação da regressão-krigagem na capacidade de generalização de Corg, esses resultados mostram que a krigagem dos resíduos da RLM não contribuiu para melhoria da predição de Corg.

### CONCLUSÕES

O software SAGA permite processar e visualizar os dados e resultados do curso/projeto, além de gerar facilmente derivadas de relevo. O software R é bastante flexível para solucionar diferentes problemas em MDS e sua implementação via RStudio facilita a criação e entendimento das rotinas e comandos utilizados. Portanto, recomendamos a utilização desses softwares em cursos/projetos de MDS.

O método de regressão-krigagem engloba diversos temas importantes em MDS, constituindo uma abordagem teórico-prática interessante para utilização em cursos nessa área.

### AGRADECIMENTOS

Agradecemos à Organização das Nações Unidas para a Alimentação e a Agricultura pelo apoio financeiro e aos participantes do curso pela paciência e atenção dedicadas.

Agradecemos também à equipe da Embrapa Solos, especialmente à Renata Paldês, Paula Almeida, Murilo Matheus, Mário Áglio e Maria Aparecida Guedes, pelo apoio.

### REFERÊNCIAS

FONTANA, A.; PEREIRA, M. G.; NASCIMENTO, G. B. et al. Matéria orgânica em solos de tabuleiros na região Norte Fluminense, RJ. *Floresta e Ambiente*, 8:114-119, 2001.

MENDONÇA-SANTOS, M. L.; SANTOS, H. G.; DART, R. O. et al. Modelagem e mapeamento digital de estoque de carbono orgânico na camada superficial dos solos (0-10 cm) do estado do Rio de Janeiro. *Boletim de Pesquisa e Desenvolvimento*, 120, 2007.

R CORE TEAM. R: a language and environment for statistical computing. Viena: R Foundation for Statistical Computing. Disponível em: <<http://cran.r-project.org>>. Acesso em 10 out. 2012.

RSTUDIO. RStudio: integrated development environment for R. Boston: RStudio. Disponível em: <<http://www.rstudio.com>>. Acesso em 10 out. 2012.

SAGA USER GROUP ASSOCIATION. System for Automated Geoscientific Analyses (SAGA). Göttingen: Department for Physical Geography, University of Göttingen. Disponível em: <<http://www.saga-gis.org>>. Acesso em 10 out. 2012.