

**RENATO CORRÊA TAQUES**

**MODELOS PEDOMÉTRICOS PARA ESPACIALIZAÇÃO DE  
ATRIBUTOS DE SOLOS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS - BRASIL  
2014

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

T175m  
2014 Taques, Renato Corrêa, 1978-  
Modelos pedométricos para espacialização de atributos de solos / Renato Corrêa Taques. – Viçosa, MG, 2014.  
vii, 68f. : il. (algumas color.) ; 29 cm.

Orientador: Elpídio Inácio Fernandes Filho.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. Mapeamento do solo. 2. Mapeamento digital.  
3. Geoprocessamento. 4. Mineração de dados (Computação).  
5. Análise espacial (Estatística). I, Universidade Federal de Viçosa. Departamento de Solos. Programa de pós-graduação em Solos e Nutrição de Plantas. II. Título.


CDD 22 ed. 631.4

RENATO CORRÊA TAQUES

MODELOS PEDOMÉTRICOS PARA ESPACIALIZAÇÃO DE  
ATRIBUTOS DE SOLOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Solos e Nutrição de Plantas, para obtenção do título de *Magister Scientiae*.

APROVADO: 04 de dezembro de 2014.



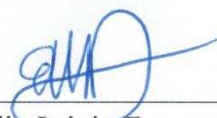
Eliana de Souza



Márcio Rocha Francelino



Nerilson Terra Santos  
(Coorientador)



Elpídio Inácio Fernandes Filho  
(Orientador)

## AGRADECIMENTOS

À Deus.

À Universidade Federal de Viçosa, pela graduação em Engenharia de Agrimensura e mestrado em Solos e Nutrição de Plantas (Ciências do Solo).

Ao Incaper (Instituto Capixaba de Pesquisa Assistência Técnica e Extensão Rural), pela liberação para me dedicar em tempo integral ao curso de mestrado.

Ao professor Elpídio Inácio Fernandes Filho, pela oportunidade, incentivo, ensinamentos e orientação tranquila e segura, que muito contribuíram para a realização deste trabalho.

Ao professor Nerilson Terra Santos, disponibilidade e coorientação, principalmente na área de Estatística Espacial.

À todos os professores do Departamento de Solos, pelo conhecimento transmitido nas disciplinas.

À todos os servidores do DPS que estão sempre cuidando para que os nossos dias sejam mais fáceis.

À FAPES (Fundação de Amparo à Pesquisa do estado de Espírito Santo) pelo apoio financeiro.

Ao Serviço Geológico Brasileiro (CPRM) pela disponibilização de dados.

À todos os colegas e amigos do Labgeo e da Pós-Graduação pelo apoio.

Aos amigos Bruno Nery, Leo Jackson, José Aridiano e Lucas Gomes, pelas experiências compartilhadas e apoio em momentos importantes.

À todos os colegas e amigos do Incaper, em especial ao amigo José Sergio pelo incentivo e apoio.

À minha família e amigos pela motivação e apoio.

Aos meus pais, José Leonir (in memoriam) e Cleusa Maria, pela vida, educação e incentivo ao estudo.

À minha amada esposa Lígia, pelo carinho, dedicação, paciência e incentivo.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Muito obrigado!

## SUMÁRIO

RESUMO .....	vi
ABSTRACT .....	vii
INTRODUÇÃO GERAL .....	1
REFERÊNCIAS BIBLIOGRÁFICAS .....	3
CAPÍTULO 1: AVALIAÇÃO DA VARIABILIDADE ESPACIAL DE ATRIBUTOS DE SOLOS .....	5
RESUMO .....	5
ABSTRACT .....	5
1.1 INTRODUÇÃO .....	6
1.2 MATERIAL E MÉTODOS .....	7
1.2.1 Área de estudo .....	7
1.2.2 Base de dados espaciais .....	8
1.2.3 Amostragem do solo .....	8
1.2.4 Análises laboratoriais .....	9
1.2.5 Análises exploratórias dos dados .....	9
1.2.6 Semivariogramas .....	10
1.2.7 Krigagem ordinária .....	12
1.3 RESULTADOS .....	12
1.3.1 Estatística descritiva .....	12
1.3.2 Análise exploratória espacial .....	13
1.3.3 Modelagem do semivariograma empírico .....	15
1.4 DISCUSSÃO .....	18
1.5 CONCLUSÕES .....	20
REFERÊNCIAS BIBLIOGRÁFICAS .....	20
CAPÍTULO 2: SELEÇÃO DE COVARIÁVEIS AMBIENTAIS E PREDIÇÃO ESPACIAL DE ATRIBUTOS DE SOLOS USANDO REGRESSÃO LINEAR MÚLTIPLA E <i>RANDOM FOREST</i> .....	22
RESUMO .....	22

ABSTRACT.....	22
2.1 INTRODUÇÃO .....	23
2.2 MATERIAL E MÉTODOS .....	24
2.2.1 Área de estudo.....	24
2.2.2 Base de dados dos atributos de solos .....	25
2.2.3 Covariáveis ambientais .....	26
2.2.4 Correlação entre atributos de solo e covariáveis ambientais .....	28
2.2.5 Seleção do MDE .....	28
2.2.6 Modelos preditivos.....	28
2.2.6.1 Regressão Linear Múltipla .....	28
2.2.6.2 <i>Random Forest</i> .....	29
2.2.6.3 Avaliação da exatidão dos modelos preditivos .....	30
2.3 RESULTADOS .....	31
2.3.1 Estatística descritiva.....	31
2.3.2 Análise da correção linear .....	31
2.3.3 Seleção de covariáveis ambientais para regressão linear múltipla.....	34
2.3.4 Importância das covariáveis ambientais nos modelos <i>Random Forest</i> .....	35
2.3.5 Mapa digital de atributos de solos e exatidão dos modelos preditivos .....	36
2.4 DISCUSSÃO .....	38
2.5 CONCLUSÕES .....	40
REFERÊNCIAS BIBLIOGRÁFICAS.....	40
CAPÍTULO 3: COMPARAÇÃO DE MODELOS GEOESTATÍSTICOS PARA A PREDIÇÃO DOS ATRIBUTOS DE SOLOS .....	46
RESUMO .....	46
ABSTRACT.....	46
3.1 INTRODUÇÃO .....	47
3.2 MATERIAL E MÉTODOS .....	48
3.2.1 Área de estudo.....	48
3.2.2 Amostragem e análises de solos.....	49

3.2.3 Covariáveis ambientais .....	49
3.2.4 Modelos preditivos.....	51
3.2.4.1 Krigagem ordinária .....	51
3.2.4.2 Cokrigagem.....	52
3.2.4.3 Regressão-krigagem.....	52
3.2.5 Validação cruzada (autovalidação) .....	53
3.3 RESULTADOS.....	54
3.3.1 Estatística descritiva dos dados dos atributos de solos .....	54
3.3.2 Covariáveis preditoras.....	55
3.3.3 Predição espacial dos atributos de solos .....	56
3.4 DISCUSSÃO .....	64
3.5 CONCLUSÕES .....	65
REFERÊNCIAS BIBLIOGRÁFICAS.....	65
CONCLUSÕES GERAIS .....	68

## RESUMO

TAQUES, Renato Corrêa, M.Sc., Universidade Federal de Viçosa, dezembro de 2014. **Modelos pedométricos para espacialização de atributos de solos.** Orientador: Elpídio Inácio Fernandes Filho. Coorientador: Nerilson Terra Santos.

Informações sobre os solos são cada vez mais requeridas para auxiliar a tomada de decisão em uma série de questões globais, tornando crescente a procura por informações atualizadas e precisas sobre os solos. Métodos geoestatísticos apresentam bons resultados em levantamentos quantitativos de solos de níveis detalhados (escalas  $\geq 1:20.000$ ) e métodos convencionais, aparentemente, são mais eficientes em levantamentos de reconhecimento de baixa intensidade (escalas  $\leq 1:250.000$ ). O desafio é encontrar métodos de levantamento quantitativo, aplicáveis em escalas intermediárias. Este trabalho teve como objetivo realizar estudos comparativos das metodologias de mapeamento digital de atributos de solos, compatível com escalas de mapeamento semidetalhado. O estudo foi realizado na bacia do Rio Pongal, no município de Anchieta/ES (Lat 20°45'S, Long 40°45'W), com área aproximada de 14.000 hectares. A base de dados de atributos de solos foi obtida a partir dos resultados de análises de rotina para determinação do pH, CTC, MOS, teores de areia, silte e argila, de 154 amostras da camada 15 a 30 cm de profundidade. Covariáveis ambientais, geradas a partir de MDEs, dados gamaespectrométricos e imagens do satélite Landsat 8, foram usadas nos modelos preditivos. O capítulo 1 é dedicado ao estudo da variabilidade espacial dos atributos de solos, com a modelagem dos semivariogramas, avaliação do alcance e da dependência espacial. Com exceção do pH, todas as variáveis apresentaram forte dependência espacial, com alcance variando entre 297 e 565 m. O capítulo 2 trata da seleção de covariáveis mais relevantes para a predição espacial, usando Regressão Linear Múltipla (RLM) e *Random Forest* (RF). As covariáveis derivadas do MDE de 5 m de resolução apresentaram os melhores resultados em relação às covariáveis dos demais MDEs estudados. Os modelos RF apresentaram resultados melhores de R<sup>2</sup> e RMSE, se comparados com os resultados da RLM. O desempenho dos modelos geoestatísticos de krigagem ordinária, cokrigagem e regressão-krigagem na predição espacial de atributos de solos foi avaliado no capítulo 3. Na avaliação final, a regressão-krigagem apresentou os melhores resultados na predição dos atributos de solos, entre os modelos geoestatísticos avaliados. Concluímos que o uso de técnicas híbridas, de regressão combinada com a krigagem, que utilizam covariáveis ambientais na predição, produzem os melhores resultados no mapeamento digital de atributos de solos em escala semidetalhadas.

## ABSTRACT

TAQUES, Renato Corrêa, M.Sc., Universidade Federal de Viçosa, December, 2014. **Pedometrics model for spatialization of soil properties**. Adviser: Elpídio Inácio Fernandes Filho. Co-adviser: Nerilson Terra Santos.

Soils information is increasingly required to aid decision taking on a range of global issues, becoming growing demand for updated and accurate information on soils. Geostatistical methods have good results in quantitative soil surveys of detailed levels (scale  $\geq 1:20,000$ ) and conventional methods are, apparently, more efficient at low intensity reconnaissance surveys ( $\leq$  scales  $1:250,000$ ). The challenge is to find quantitative survey methods, applicable to intermediate scales. This work aimed to conduct comparative studies of methodologies of digital mapping of soil properties compatible with semi-detailed scales mapping. The study was conducted at the Rio Pongal basin, in the municipality of Anchieta/ES (Lat  $20^{\circ}45'S$ ; Long  $40^{\circ}45'W$ ), with an approximate area of 14,000 hectares. The database of soil properties was obtained from the results of routine analysis to determine pH, CEC, SOM, sand, silt and clay, 154 samples layer of 15 to 30 cm deep. Environmental covariates, generated from DEMs, gamma ray spectrometric data and images from the satellite Landsat 8, were used in predictive models. Chapter 1 is devoted to the study of the spatial variability of soil properties, with modeling the semivariogram, evaluation value range and spatial dependence. With exception of pH, all variables showed a strong spatial dependence, with range varying between 297 and 565 m. Chapter 2 deals with the selection of the most relevant variables for the spatial prediction using Multiple Linear Regression (MLR) and Random Forest (RF). The covariates derived from the DEM of 5 m resolution showed the best results in relation to other DEMs studied. RF models showed the best results in  $R^2$  and RMSE, compared with the results of MLR. The performance of geostatistical models of ordinary kriging, cokriging and regression-kriging in spatial prediction of soil properties was evaluated in Chapter 3. In the final evaluation, the regression-kriging showed the best results in the prediction of soil properties, between models geostatistical evaluated. We conclude that using hybrid techniques, regression combined with the kriging, using environmental covariates to predict, produce the best results in digital mapping of soil properties in semi-detailed scales.

## INTRODUÇÃO GERAL

Informações sobre os solos são cada vez mais requeridas para auxiliar a tomada de decisão em uma série de questões globais, tais como: a produção de alimentos e a erradicação da fome, mudanças climáticas e degradação ambiental, tornando crescente a procura por informações atualizadas e precisas. A comunicação dessas informações, de forma adequada à complexidade dos solos na paisagem e, ao mesmo tempo, de fácil compreensão, junto com a atualização dos métodos utilizados para melhoria da exatidão, ainda é um desafio (Sanchez *et al.*, 2009).

Métodos geoestatísticos apresentam bons resultados em levantamentos quantitativos do solo de níveis detalhados (escalas  $\geq 1:20.000$ ). Por outro lado, métodos convencionais, aparentemente, são mais eficientes em levantamentos de reconhecimento de baixa intensidade (escalas  $\leq 1:250.000$ ). Existe, no entanto, uma carência de informações sobre métodos de levantamento quantitativo aplicáveis em escalas intermediárias (ex: 1:50.000 e 1:100.000). Nestas escalas, a dificuldade para o mapeamento preditivo de atributos de solos se dá tanto pelo tamanho da área, como por sua complexidade geomorfológica (McKenzie e Ryan, 1999).

Predição espacial de atributos de solos utilizando covariáveis ambientais (auxiliares), espacialmente densas, tem atraído muitas pesquisas na área de pedometria. McBratney *et al.* (2003) destaca que a maioria das pesquisas em pedometria tenta usar as informações auxiliares como covariáveis, e que estas covariáveis podem estar relacionadas à formação do solo ou aos fatores SCORPAN, que representam: o próprio solo (S), o clima (C), organismos vivos (O), relevo (R), material de origem (P), o tempo (A) e as coordenadas geográficas (N).

Neste contexto, o uso de covariáveis ambientais, pelos modelos preditivos, visa à melhoria da exatidão das predições. Entre as covariáveis ambientais mais promissoras para a predição espacial de atributos de solos, merecem destaque os atributos de terreno (Florinsky *et al.*, 2002), as covariáveis derivadas de sensoriamento remoto do espectro visível ao infravermelho (Lu *et al.*, 2013) e de gamaespectrometria (Wilford e Minasny, 2007).

Os atributos de terreno revelam os processos geomorfológicos, climáticos e hidrológicos. Entretanto, Modelo Digital de Elevação (MDE) de diferentes resoluções espaciais produzem diferentes atributos de terreno. A escolha da resolução espacial depende da complexidade da paisagem e dos atributos de terreno utilizadas na modelagem (Ten Caten *et al.*, 2012).

Observações de campo e experimentos de laboratório têm mostrado que o sensoriamento remoto do espectro visível ao infravermelho próximo (vis-NIR) pode fornecer informações sobre atributos de solos, relacionados com a cor, óxido de ferro, minerais de argila, tamanho de partículas e teor de matéria orgânica do solo. Estas informações são obtidas através da absorção específica de comprimentos de onda da radiação eletromagnética na faixa de 400 a 2500 nm (Rossel e Chen, 2011).

A gamaespectrometria é uma técnica passiva de sensoriamento remoto, que se baseia na emissão terrestre de raios gama, uma forma de alta energia e curto comprimento de onda da radiação eletromagnética. A relativa abundância ou concentração desses radioelementos nos solos e rochas é estimada por meio das suas intensidades em seus picos de emitância. A espectrometria teletransportada mede a abundância de raios gama de Potássio (K), de Tório (Th) e de Urânio (U) em rochas e materiais intemperizados, por meio da detecção dos raios gama emitidos do decaimento ativo e natural dos isótopos radioativos desses elementos (Wilford *et al.*, 1997).

Cabe ressaltar que o interesse pelo desenvolvimento de métodos quantitativos de levantamento aplicado ao mapeamento de atributos de solos, em uma resolução fina (alta resolução), para áreas mais extensas, é relativamente recente. Este fato é confirmado pelo pequeno número de trabalhos realizados e publicados com este tema, a exemplo dos estudos feitos por Henderson *et al.* (2005), Li (2010), Sun *et al.* (2012), Lu *et al.* (2013), Souza *et al.* (2014), Carvalho Junior *et al.* (2014) e Qi-yong *et al.* (2014).

Entre os atributos de solos previstos para o mapeamento digital de solos (GlobalSoilMap.net, 2011), o carbono orgânico, pH em água, teor de argila, teor de silte, teor de areia e a capacidade de troca catiônica, são considerados primários, e merecem destaque por terem um papel muito importante nos fenômenos físico-químicos que ocorrem no solo. Os atributos primários podem ser relacionados à área específica do solo, retenção de água, estoque de carbono, fertilidade, fixação de fósforo, textura, densidade do solo, suscetibilidade à erosão, entre outras.

Diante do exposto, este trabalho teve como objetivo realizar estudos comparativos das metodologias de mapeamento digital de atributos de solos, aplicado a áreas extensas, com resolução fina, compatível com escalas de mapeamento semidetalhado e de reconhecimento de alta intensidade. O capítulo 1 é dedicado a estudar a variabilidade espacial dos atributos de solos, e estimar a densidade de amostras necessárias para predição espacial dos atributos em escala de mapeamento semidetalhado, usando a geoestatística. O capítulo 2 trata da seleção de covariáveis ambientais mais relevantes para a predição espacial e da predição dos atributos de solos

com modelos de Regressão Linear Múltipla (RLM) e *Random Forest* (RF). O desempenho dos modelos de predição espacial de atributos de solos, com a incorporação de informações de covariáveis ambientais, foi avaliado no capítulo 3.

Os resultados obtidos no estudo poderão respaldar metodologias para a predição espacial de atributos de solos em escala regional, e serem usadas como referência para estudos similares, em outras regiões, que busquem reduzir o tempo e o investimento necessários para a obtenção de informações espacialmente referenciadas, precisas e atualizadas sobre solos.

## REFERÊNCIAS BIBLIOGRÁFICAS

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. 232-234, p. 479-486, 2014.

FLORINSKY, I. V.; EILERS, R. G.; MANNING, G. R.; FULLER, L. G. Prediction of soil properties by digital terrain modelling. **Environmental Modelling & Software**, v. 17, p. 295-311, 2002.

GLOBALSOILMAP.NET. **Specifications Version 1 GlobalSoilMap.net products**. Disponível em: < <http://globalsoilmap.net/> > Acesso em: julho de 2013.

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, n. 3-4, p. 383-398, 2005.

LI, Y. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? **Geoderma**, v. 159, n. 1-2, p. 63-75, 2010.

LU, P.; WANG, L.; NIU, Z.; LI, L.; ZHANG, W. Prediction of soil properties using laboratory VIS-NIR spectroscopy and Hyperion imagery. **Journal of Geochemical Exploration**, v. 132, p. 26-33, 2013.

MCBRATNEY, A. B.; SANTOS, M. L. M.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, p. 3-52, 2003.

MCKENZIE, N. J.; RYAN, P. J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, v. 89, n. 1-2, p. 67-94, 1999.

MINASNY, B.; MCBRATNEY, A. B. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. **Geoderma**, v. 140, p. 324-336, 2007.

PRACILIO, G.; ADAMS, M. L.; SMETTEM, K. R. J.; HARPER, R. J. Determination of spatial distribution patterns of clay and plant available potassium contents in surface soils at the farm scale using high resolution gamma ray spectrometry. **Plant and Soil**, v. 282, n. 1-2, p. 67-82, 2006.

QI-YONG, Y.; ZHONG-CHENG, J.; WEN-JUN, L.; HUI, L. Prediction of soil organic matter in peak-cluster depression region using kriging and terrain indices. **Soil & Tillage Research**, v. 144, p. 126-132, 2014.

ROSSEL, R. A. V.; CHEN, C. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. **Remote Sensing of Environment**, v. 115, n. 6, p. 1443-1455, 2011.

SANCHEZ, P. A.; AHAMED, S.; CARRE, F.; HARTEMINK, A. E.; HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; MCBRATNEY, A. B.; MCKENZIE, N. J.; MENDONÇA-SANTOS, M. L.; MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C. A.; SACHS, J. D.; SHEPHERD, K. D.; VAGEN, T. G.; VANLAUWE, B.; WALSH, M. G.; WINOWIECKI, L. A.; ZHANG, G. L. Digital soil map of the world. **Science**, v. 325, n. 5941, p. 680-681, 2009.

SOUZA, E.; HENGL, T.; KEMPEN, B.; HEUVELINK, G. B. M.; FERNANDES FILHO, E. I.; SCHAEFER, C. E. G. R. Comparing spatial prediction methods for soil property mapping in Brazil: A case study for the Rio Doce Basin. In: ARROUAYS, D.; MCKENZIE, N. J., *et al.* (Eds.). **GlobalSoilMap - Basis of the global spatial soil information system** ed. Leiden, Holanda: CRC Press, 2014. p. 267-271.

SUN, W.; MINASNY, B.; MCBRATNEY, A. Analysis and prediction of soil properties using local regression-kriging. **Geoderma**, v. 171-172, p. 16-23, 2012.

TEN CATEN, A.; DALMOLIN, R. S. D.; PEDRON, F. D. A.; MENDONÇA-SANTOS, M. L. Resolução espacial de um modelo digital de elevação definida pela função wavelet. **Pesq. agropec. bras.**, v. 47, n. 3, p. 449-457, 2012.

WILFORD, J. R.; BIERWIRTH, P. N.; CRAIG, M. A. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. **Journal of Australian Geology & Geophysics**, v. 17, p. 201 - 216, 1997.

WILFORD, J. R.; MINASNY, B. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. In: LAGACHERIE, P.; MCBRATNEY, A. B., *et al.* (Eds.). **Digital Soil Mapping: An Introductory Perspective** ed. Amsterdam: Elsevier, 2007. p. 207-218. (Developments in Soil Science, 31).

# **CAPÍTULO 1: AVALIAÇÃO DA VARIABILIDADE ESPACIAL DE ATRIBUTOS DE SOLOS**

## **RESUMO**

O solo é um importante recurso natural, cuja gestão requer o conhecimento e informações sobre as propriedades e atributos que o caracterizam. A variabilidade espacial dos atributos de solo é uma preocupação antiga dos pesquisadores da Ciência do Solo, e o semivariograma tem um papel importante na análise espacial da variabilidade dos atributos de solo, resumindo a variação do atributo dentro de uma região. Os objetivos deste estudo foram avaliar a variabilidade espacial dos atributos físico-químicas de solo (pH, CTC, MOS e teores de areia, silte e argila) em escala semidetalhada de levantamento de solos, e estimar a densidade amostral necessárias para a predição espacial dos atributos de solos. O estudo foi realizado na bacia do Rio Pongal, localizada na região litorânea sul do estado do Espírito Santo, com aproximadamente 14.000 hectares. Os resultados mostram que: o pH apresenta maior variabilidade espacial entre os atributos de solos, exigindo a maior densidade amostral para sua predição em escalas semidetalhadas; a variabilidade espacial da CTC, MOS, areia, silte e argila, apresenta forte dependência espacial, permitindo o uso de modelos geoestatísticos para predições; e que o uso de técnicas geoestatísticas é viável para obtenção de informações espaciais sobre a variabilidade dos atributos de solos em escalas semidetalhadas de levantamentos.

Termos para indexação: semivariograma, dependência espacial, densidade de amostras, krigagem ordinária.

## **ASSESSMENT OF SPATIAL VARIABILITY OF SOIL PROPERTIES**

### **ABSTRACT**

Soil is an important natural resource, and its management requires knowledge and information about the properties and attributes that characterize it. The spatial variability of soil properties is an old concern of the researchers of Soil Science, and semivariogram has an important role in spatial analysis of the variability of soil properties, summarizing the variation of the property within a region. The objectives of this study were to evaluate the spatial variability of physical and chemical soil properties (pH, CEC, SOM, sand, silt and clay) in semi-detailed scale soil survey and

estimate the sample density required for the spatial prediction of soil properties. The study was conducted at the Rio Pongal basin, located in the south coastal region the state of Espírito Santo, with approximately 14,000 hectares. The results show that: the pH presents the more spatial variability between the properties of soils, requiring larger sample density for his prediction in semi-detailed scales; the spatial variability of CEC, SOM, sand, silt and clay, has a strong spatial dependence, allowing the use of geostatistical models for predictions; and the use of geostatistical techniques is viable to obtain spatial information on the variability of soil properties in semi-detailed scale surveys.

Index terms: semivariogram, spatial dependence, density samples, ordinary kriging.

## 1.1 INTRODUÇÃO

Os recursos naturais da terra não são inesgotáveis, e a crescente demanda imposta ultrapassa a capacidade natural de repor tais recursos. O solo é um componente importante dos recursos naturais, e sua gestão requer o conhecimento e informações sobre as propriedades e atributos que o caracteriza (Carvalho Junior *et al.*, 2014).

A variabilidade espacial dos atributos do solo é uma preocupação antiga dos pesquisadores da Ciência do Solo (Vieira, 2000). A incorporação da componente espacial nos estudos dos solos é uma forma mais adequada para descrever o comportamento dos seus atributos. Esta abordagem permite avaliar a variabilidade espacial dos atributos de solo, bem como representá-las na forma de mapas.

O semivariograma tem um papel importante na análise espacial da variabilidade dos atributos de solo, resumindo a variação de um atributo dentro de uma região (Webster e Oliver, 1992). A partir do semivariograma é possível avaliar a existência de dependência espacial dos dados, que é essencial para aplicação de técnicas geoestatísticas como a krigagem e a cokrigagem.

As técnicas geoestatísticas apresentam grande potencial em levantamentos semidetalhados de solo (ex: escala 1/50.000). Sua aplicação ainda é restrita, devido à densidade de amostras necessárias para ajustar os modelos, que têm relação direta com o custo e o tempo de execução dos levantamentos. A estimativa da densidade de amostras permite avaliar a viabilidade técnica e econômica da aplicação das técnicas geoestatística para o mapeamento dos atributos do solo.

Os objetivos deste estudo foram avaliar a variabilidade espacial dos atributos físico-químicos do solo (pH, CTC, MOS e teores de areia, silte e argila) em escala semidetalhada de levantamento de solos, e estimar a densidade de amostras necessárias para a predição espacial dos atributos de solos.

## 1.2 MATERIAL E MÉTODOS

### 1.2.1 Área de estudo

O estudo foi realizado para a bacia do Rio Pongal, localizada na região litorânea sul do estado do Espírito Santo, na zona rural do município de Anchieta. A área possui aproximadamente 14.000 hectares e localiza-se entre as latitudes 20°41'S a 20°48'S, e longitudes 40°38'W a 40°49'W (Figura 1).

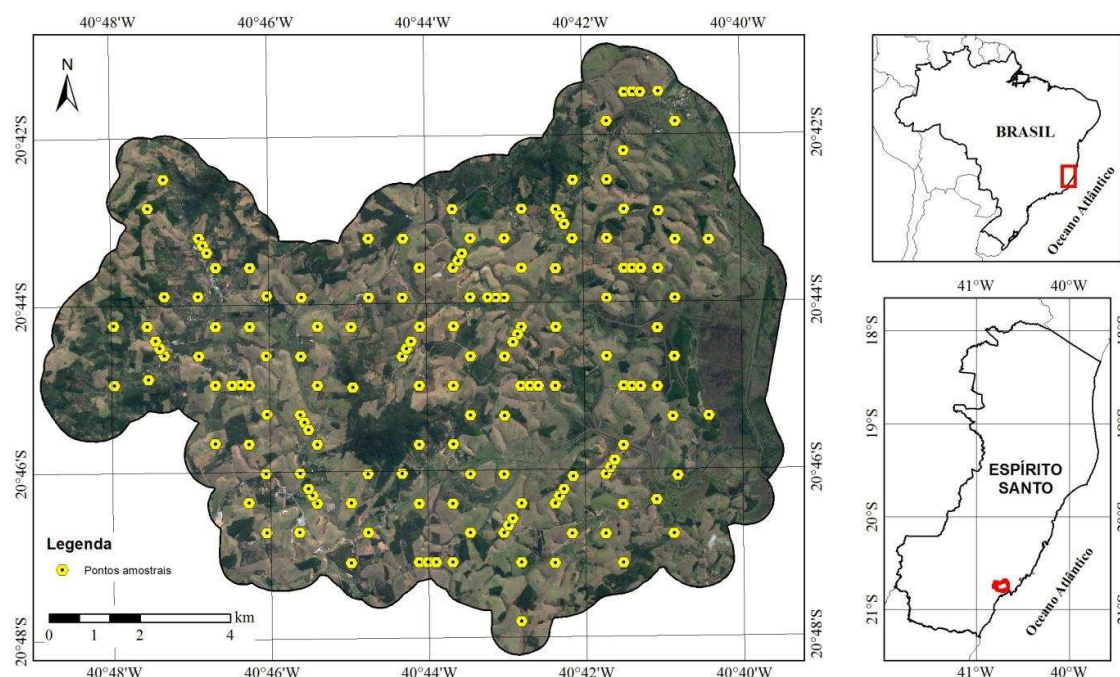


Figura 1 - Área de estudo e localização dos pontos de amostragem de solo.

A bacia hidrográfica apresenta altitudes entre 0 e 600 m, com relevo variando de plano, localizado nas partes baixas, até escarpado, nas partes mais elevada. Relevos suave ondulado, ondulado, forte ondulado e montanhoso ocorrem em posições intermediárias da bacia (IBGE, 2007). A geologia é diversificada, formada por gnaisses aluminosos, ortognaisses graníticos, ortognaisses granulíticos e também formação do grupo barreiras e depósitos sedimentares (CPRM, 1993).

A área de estudo possui pedologia complexa. As principais classes de solos que ocorrem na bacia do Rio Pongal são: Latossolos Vermelho-Amarelo Distróficos,

Cambissolos Háplico Tb Distróficos, Argissolos Amarelo Distrocoesos, Latossolos Amarelo Distróficos, Gleissolos Háplico Tb Distróficos, Gleissolos Háplico Sódicos e Neossolos Flúvico Tb Distróficos (RADAMBRASIL, 1983).

### **1.2.2 Base de dados espaciais**

Foram usadas ferramentas computacionais específicas de geoprocessamento, aplicadas no desenvolvimento do Sistema de Informações Geográficas (SIG), com uso do software ArcGIS Desktop versão 10.2 (ESRI, 2013). Todos os dados utilizados no desenvolvimento do SIG foram armazenados em um banco de dados geográfico (geodatabase), tendo como datum SIRGAS 2000 e o sistema de projeção Universal Transversa de Mercator (UTM).

A base de dados espacial foi construída com dados do mapa de solos do levantamento original do Projeto RADAMBRASIL, em escala 1:250.000 (RADAMBRASIL, 1983), carta geológica do Programa de Levantamento Geológico Básico do Brasil, em escala 1:100.000 (CPRM, 1993), dados do relevo (hipsometria, pontos de elevação, modelos digitais de elevação), hidrografia e imagens de sensoriamento remoto (orbital e aerotransportados).

### **1.2.3 Amostragem do solo**

A localização dos pontos de coleta de amostras foi definida a partir de uma grade hexagonal regular, com 750 m de lado (Yfantis *et al.*, 1987). A grade hexagonal foi gerada pelo aplicativo “Repeating Shapes for ArcGIS”, selecionando o limite da bacia hidrográfica para indicar a extensão da grade. Foram coletados 118 pontos da grade hexagonal básica.

Um segundo conjunto, contendo 36 pontos amostrais, foi coletado para avaliação da variabilidade espacial dos dados em intervalos de distâncias menores que 750 m (Goovaerts, 1997). Para isso, foram selecionados 18 vértices (pontos já amostrados) da grade regular básica, distribuídos aleatoriamente dentro da área de estudo. Partindo de cada um dos vértices selecionados, e alinhados à direção de um lado do hexágono original, foi estabelecido um ponto a 187 m e outro a 375 m de distância do vértice.

Em campo, a localização do ponto de coleta foi determinada com receptor GPS de navegação, o mais próximo possível da posição determinada previamente. Após a localização do ponto, a posição de coleta da amostra foi georreferenciada com posicionamento diferencial (DGPS).

Foram abertas minitrincheiras para coleta das amostras de solo na camada entre 15 e 30 cm de profundidade, de acordo com as especificações da GlobalSoilMap.net (2011), totalizando 154 amostras coletadas. As amostras foram acondicionadas em sacos plásticos adequados ao transporte e armazenamento, devidamente identificados com etiquetas. Foram realizadas observações complementares para caracterização morfológica das amostras (Santos *et al.*, 2013).

#### **1.2.4 Análises laboratoriais**

As amostras de solo foram peneiradas em malha de 2 mm, após sua completa dessecação ao ar, para obtenção da terra fina seca ao ar (TFSA), e, em seguida, encaminhadas para análises laboratoriais de rotina.

As seguintes análises laboratoriais foram realizadas para a caracterização química das amostras, seguindo os métodos constantes em Embrapa (2011): pH H<sub>2</sub>O (-log [H<sup>+</sup>]), com a medição eletrônica do potencial por meio de eletrodo combinado imerso em suspensão 1:2,5 solo-água; CTC (capacidade de troca catiônica a pH 7 em cmol<sub>c</sub> dm<sup>-3</sup>), determinada pela soma de bases (Ca<sup>2+</sup>+Mg<sup>2+</sup>+K<sup>+</sup>) mais acidez potencial (H<sup>+</sup>+Al<sup>3+</sup>); e MOS (matéria orgânica do solo em dag kg<sup>-1</sup>), obtida por um método indireto, no qual o carbono presente na MOS é oxidado para CO<sub>2</sub> e a massa de C é determinada de acordo com a quantidade de agente oxidante (dicromato) excedente.

A caracterização física foi realizada com análise granulométrica (dispersão total), para a determinação dos teores de areia, silte e argila em dag kg<sup>-1</sup>. A análise é realizada a partir 10 g de TFSA, que passa por pré-tratamento para remoção da matéria orgânica, utilizando uma solução de NaOH 0,1 mol L<sup>-1</sup>. Posteriormente, a solução é colocada para agitar, a uma rotação de 50 rpm por 16 horas. A fração areia é separada por peneira de malha de 0,053 mm, e a fração argila e silte são determinadas por sedimentação diferencial (Ruiz, 2005; Embrapa, 2011).

Os resultados das análises laboratoriais foram tabulados, com as respectivas localizações geográficas (coordenadas UTM fuso 24 sul, datum SIRGAS 2000), em formato adequado para integrar o banco de dados espacial dentro do ambiente SIG.

#### **1.2.5 Análises exploratórias dos dados**

Para estas análises foi usado o software estatístico R (R DEVELOPMENT CORE TEAM, 2013), com os pacotes “geoR” e “outliers”. Estas análises buscam identificar e avaliar os valores discrepantes (*outliers*), pois eles têm grande influência sobre o semivariograma, bem como a existência de tendência nos dados.

Primeiro foram identificados, avaliados e removidos os valores discrepantes (*outliers*) do conjunto de dados, utilizando o critério “Z-Scores” maiores que três. Em seguida, foram calculadas estatísticas descritivas (valor mínimo, máximo, média, mediana, variância, desvio padrão, coeficiente de variação, coeficiente de assimetria e curtose) para cada variável (pH, CTC, MOS e teores de areia, silte e argila).

Também foram realizadas análises exploratórias espaciais, a partir dos mapas de distribuição espacial de quartis de cada atributo de solo. O objetivo dessas análises é avaliar a existência de tendência espacial dos dados, verificando a existência de concentração de valores de um quartil em uma dada região do espaço.

### 1.2.6 Semivariogramas

O semivariograma foi usado para medir a dependência espacial das variáveis em estudo. A semivariância é definida como a metade da variância da diferença entre as variáveis aleatórias regionalizadas,  $Z(x)$  e  $Z(x + h)$ , separadas por um vetor  $\mathbf{h}$ , conforme equação (1):

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x) - Z(x + h)] \quad (1)$$

e pode ser estimado pela equação (2):

$$\gamma(h)_{exp} = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x + h)]^2 \quad (2)$$

em que  $N(h)$  é o número de pares de valores  $Z(x)$  e  $Z(x + h)$ , separados por um vetor  $\mathbf{h}$  (Pereira *et al.*, 2013). O gráfico de  $\gamma(h)_{exp}$  em função da distância  $\mathbf{h}$  é chamado semivariograma experimental.

Inicialmente foi considerada a isotropia, quando o semivariograma é idêntico para qualquer direção de  $\mathbf{h}$  (semivariograma omnidirecional), que permite maior número de pareamentos entre pontos que os semivariogramas direcionais. Em seguida, foram examinados os mapas de semivariância, que permitem uma avaliação em 360 graus da distribuição dos valores estimados para a semivariância, para verificar a existência de tendência espacial nos dados. Os mapas de semivariância foram gerados com a ferramenta “Geostatistical Wizard” da extensão “Geostatistical Analyst” do ArcGIS 10.2 (ESRI, 2013).

O modelo teórico representa uma função contínua para semivariância, e é um dos aspectos mais importantes das aplicações geoestatísticas. Os principais modelos teóricos, aplicados são:

a) Modelo esférico

$$\gamma(h)_{mod} = C0 + C1 \left[ \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right]; 0 < h < a \quad (3)$$

b) Modelo exponencial

$$\gamma(h)_{mod} = C0 + C1 \left[ 1 - \exp \left( -3 \frac{h}{a} \right) \right]; 0 < h < d \quad (4)$$

c) Modelo gaussiano

$$\gamma(h)_{mod} = C0 + C1 \left[ 1 - \exp \left( -3 \left( \frac{h}{a} \right)^2 \right) \right]; 0 < h < d \quad (5)$$

em que  $C0$  é o efeito pepita (*nugget*),  $C0+C1$  é o patamar (*sill*),  $a$  é o alcance e  $d$  é a distância máxima na qual o semivariograma é definido (Goovaerts, 1997).

Um semivariograma experimental foi gerado para cada variável, usando o pacote de geoestatística “gstat” do software R. Os ajustes dos modelos teóricos aos semivariogramas experimentais foram feitos por Mínimos Quadrados Ponderados (WLS - *Weighted Least Squares*) usando a função “fit.variogram” para obtenção dos coeficientes efeito pepita ( $C0$ ), contribuição ( $C1$ ), patamar ( $C0+C1$ ) e alcance ( $a$ ).

A dependência espacial (DE) representa a proporção da contribuição no patamar, conforme equação (6). Zimback (2001) propõe três intervalos para classificação da dependência: forte ( $DE > 0,75$ ), moderada ( $0,25 < DE < 0,75$ ) e fraca ( $DE < 0,25$ ).

$$DE = \frac{C1}{(C0+C1)} \quad (6)$$

O valor do alcance foi considerado para o cálculo da densidade de amostras requeridas para aplicação da krigagem (Equação 7).

$$Da = \frac{10000}{a^2} \quad (7)$$

em que  $Da$  é a densidade de amostras por hectare (ha) e  $a$  é o alcance em metros.

Para distâncias menores que o alcance, os valores são correlacionados e podem ser utilizados para interpolação dos locais não amostrados. O alcance também indica a linha divisória entre a aplicação da Geoestatística ou da Estatística Clássica (Vieira, 2000).

### 1.2.7 Krigagem ordinária

Após a modelagem dos semivariogramas, a krigagem ordinária (KO) foi usada para a predição dos valores dos atributos de solos em posições não amostras. A krigagem ordinária estima localmente a média, e é aplicada em situações em que não é conhecida a média estacionária em toda a área. Caracterizada pela condição de não viés, em que a soma dos pesos deve ser a um. O seu estimador é definido como:

$$V_{KO}^* = \sum_{i=1}^n \lambda_i V_i \quad (8)$$

em que  $V_i$  são os valores da variável nos pontos amostrados  $i$ ,  $\lambda_i$  são os pesos da krigagem, que dependem da estrutura de autocorrelação espacial da variável, e são determinados de modo que a variância do erro de predição seja minimizada.

Foram obtidas também as variâncias das estimativas, que é uma medida da confiança associada às estimativas. Esta é uma característica que diferencia os métodos geoestatísticos dos demais métodos de interpolação (Vieira, 2000).

## 1.3 RESULTADOS

### 1.3.1 Estatística descritiva

O coeficiente de variação da CTC, MOS, da areia e do silte é alto ( $CV > 0,30$ ), comparado com o coeficiente de variação da argila e do pH, que apresentaram os menores valores (Tabela 1).

Tabela 1 - Estatísticas descritivas dos atributos do solo, na camada 15 a 30 cm de profundidade, da bacia do Rio Pongal.

	pH	CTC	MOS	Areia	Silte	Argila
N	153	152	149	153	153	153
Máximo	6,53	21,92	7,93	88,00	25,00	84,00
Mínimo	3,63	2,58	0,52	1,00	1,00	7,00
1° quartil	4,78	5,28	1,81	36,00	6,00	42,00
3° quartil	5,32	6,86	2,71	50,00	10,00	56,00
Média	5,06	6,51	2,33	42,83	8,19	48,41
Mediana	4,97	6,10	2,20	42,00	8,00	49,00
Variância	0,26	6,92	0,82	224,16	16,50	188,83
Desvio padrão	0,51	2,63	0,91	14,97	4,06	13,74
Assimetria	0,45	3,28	2,51	-0,30	1,51	-0,36
Curtose	1,03	13,86	11,97	1,64	3,67	0,86
CV	0,10	0,40	0,39	0,35	0,50	0,28

pH: pH em água ( $-\log[H^+]$ ); CTC: capacidade de troca catiônica a pH 7 ( $\text{cmol}_c \text{dm}^{-3}$ ); MOS: matéria orgânica do solo ( $\text{dag kg}^{-1}$ ); Areia: teor de areia ( $\text{dag kg}^{-1}$ ); Silte: teor de silte ( $\text{dag kg}^{-1}$ ); Argila: teor de argila ( $\text{dag kg}^{-1}$ ); n: número de amostras; CV: coeficiente de variação.

Os resultados do teste Shapiro-Wilk indicaram que todos os atributos em análise (pH, CTC, MOS, areia, silte e argila) não seguem uma distribuição normal ao nível de significância de 1 % (p-valor < 0,01). No entanto, pH, areia e argila apresentam valores de assimetria e curtose menores que os valores obtidos para CTC, MOS e silte.

Mesmo após a remoção dos *outliers*, a CTC, a MOS e o silte apresentaram valores elevados para o coeficiente de assimetria e curtose, indicando que a maioria dos valores é menor que a média, e a distribuição é leptocúrtica.

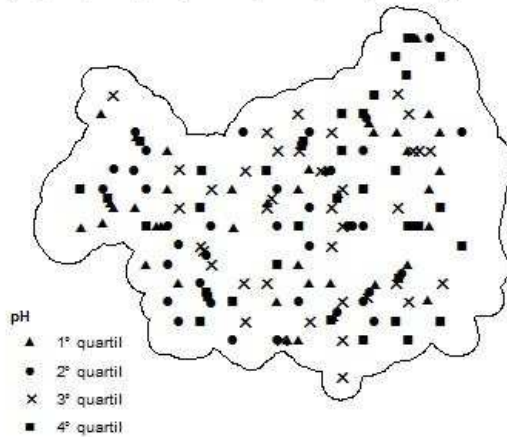
### **1.3.2 Análise exploratória espacial**

Os valores de cada quartil dos atributos de solos não apresentam um agrupamento expressivo, e estão bem distribuídos por toda a área de estudo (Figuras 2: a1, b1 e c1; Figura 3: a1, b1 e c1).

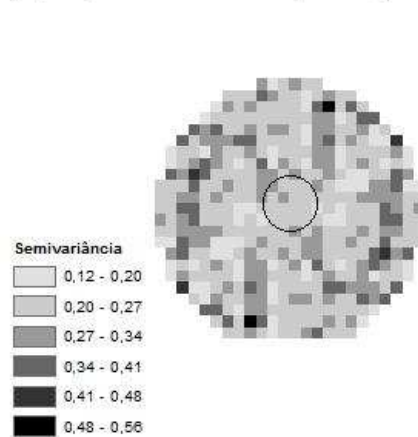
O mapa de semivariância apresenta os valores da semivariância experimental em um raio de 4500 m de distância. Os valores foram estimados para um *lag* de 375 m, que é o tamanho relativo a cada pixel da semivariância representada no mapa (Figuras 2: a2, b2 e c2; Figura 3: a2, b2 e c2).

Os mapas de semivariância indicam que a distribuição dos valores de semivariância, de cada uma das variáveis (pH, CTC, MOS, areia, silte e argila), é aproximadamente igual em todas as direções, o que caracteriza a condição de isotropia.

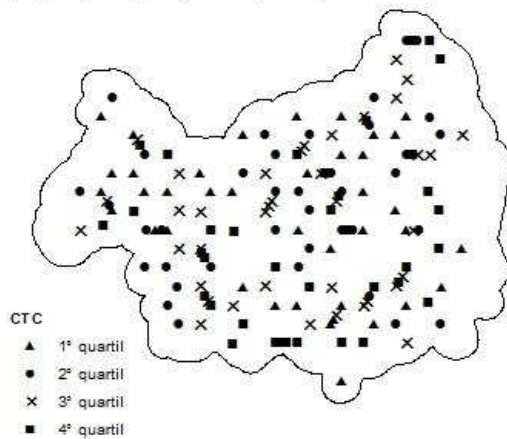
(a1) Mapa de posições de quartis para pH em água



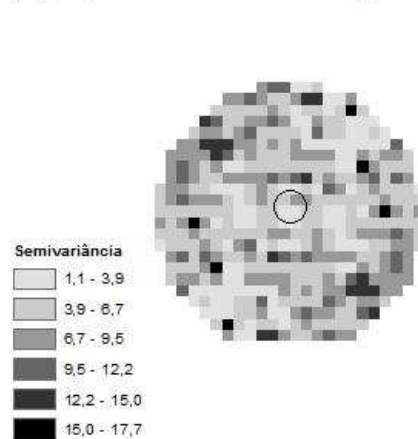
(a2) Mapa de semivariância do pH em água



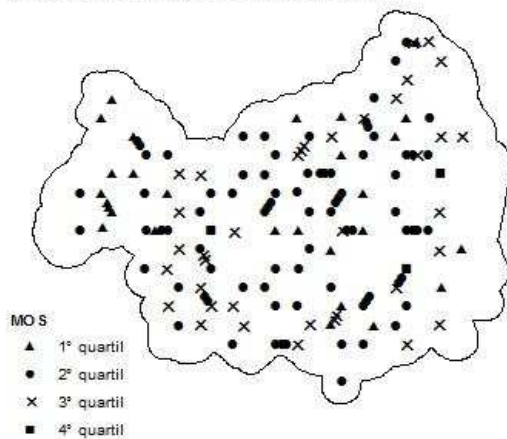
(b1) Mapa de posições de quartis para CTC a pH 7



(b2) Mapa de semivariância da CTC a pH 7



(c1) Mapa de posições de quartis para MOS



(c2) Mapa de semivariância da MOS

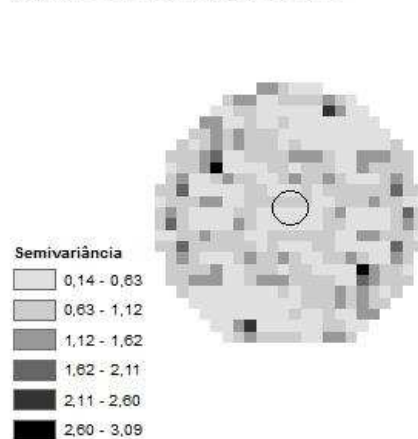
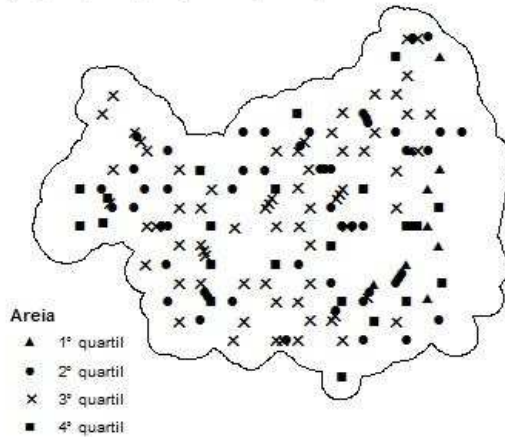
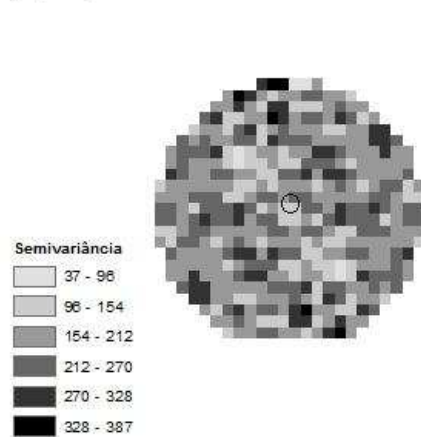


Figura 2 - Variabilidade espacial dos atributos químicos do solo, na camada de 15 a 30 cm de profundidade, da bacia do Rio Pongal.

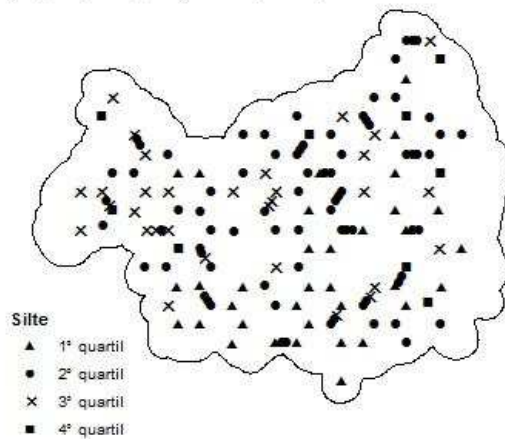
(a1) Mapa de posições de quartis para teor de areia



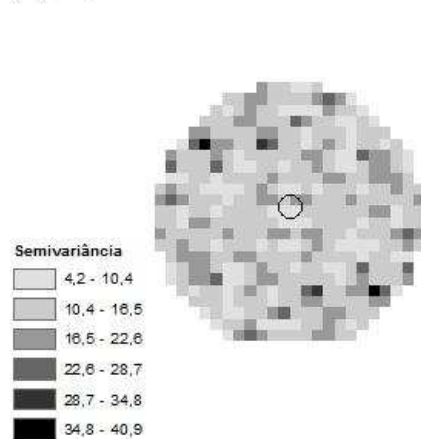
(a2) Mapa de semivariância do teor de areia



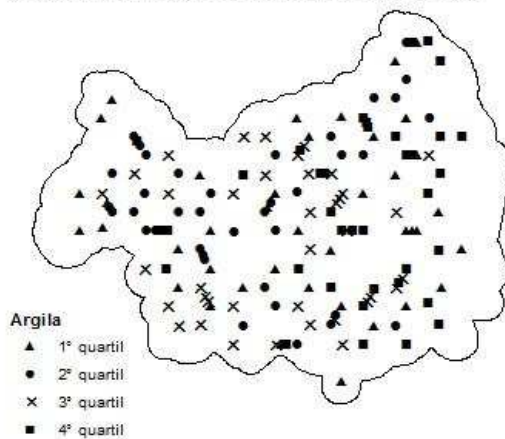
(b1) Mapa de posições de quartis para teor de silte



(b2) Mapa de semivariância do teor de silte



(c1) Mapa de posições de quartis para teor de argila



(c2) Mapa de semivariância do teor de argila

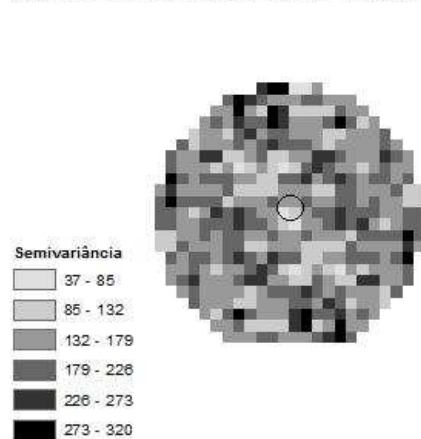


Figura 3 - Variabilidade espacial dos atributos físicos do solo, na camada de 15 a 30 cm de profundidade, da bacia do Rio Pongal.

### 1.3.3 Modelagem do semivariograma empírico

Considerando a isotropia, foram modelados os semivariogramas omnidirecionais das variáveis (Figura 4).

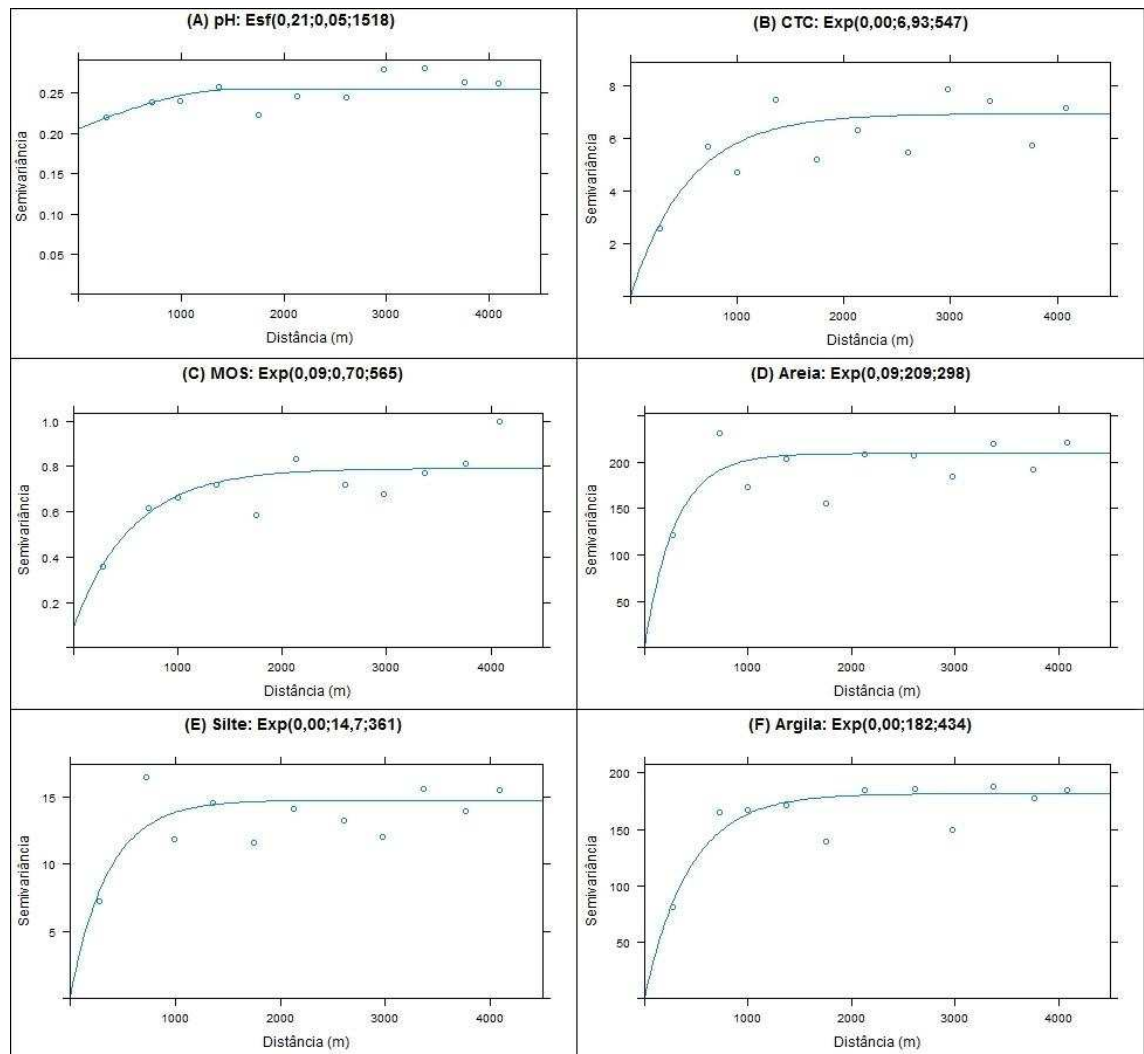


Figura 4 - Modelos teóricos ajustados aos semivariogramas empíricos das variáveis: (A) pH H<sub>2</sub>O (-log [H<sup>+</sup>]); (B) CTC (cmol<sub>c</sub> dm<sup>-3</sup>); (C) MOS (dag kg<sup>-1</sup>); (D) areia (dag kg<sup>-1</sup>); (E) silte (dag kg<sup>-1</sup>) e (F) argila (dag kg<sup>-1</sup>). Modelo (Pepita; Contribuição; Alcance). Esf - Esférico; Exp - Exponencial.

A variável pH apresentou o menor valor para a dependência espacial (DE), que indica fraca dependência espacial. As demais variáveis analisadas apresentaram forte dependência espacial (DE < 0,75), com alcance do modelo ajustado do semivariograma variando aproximadamente entre 300 e 565 m para o modelo exponencial (Tabela 2).

As variáveis que apresentaram forte dependência espacial foram interpoladas pelo método da krigagem ordinária (Figura 5). As superfícies interpoladas mostram a distribuição espacial das variáveis, o grau de homogeneidade da distribuição e permite avaliar a existência de padrões.

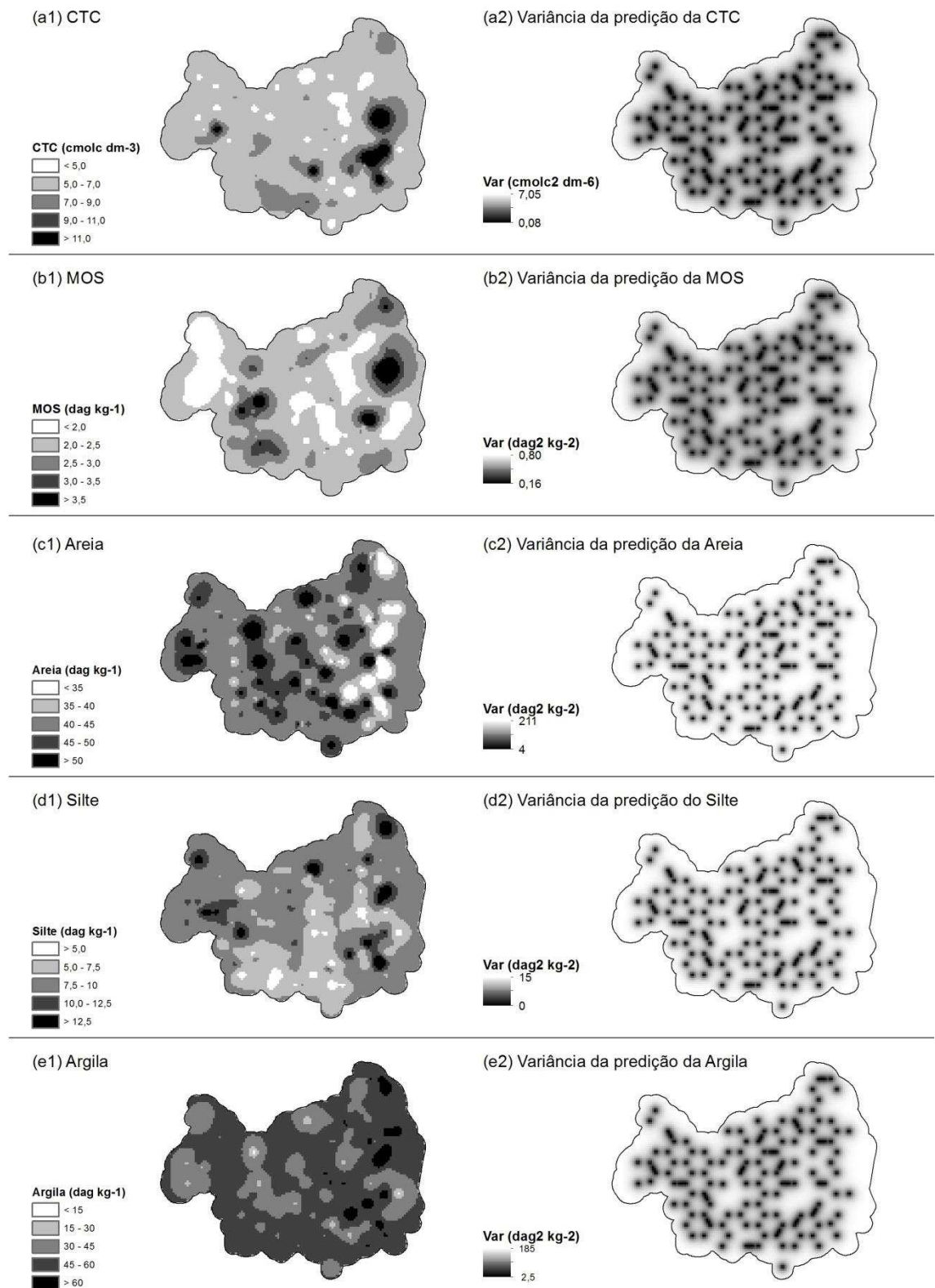


Figura 5 - Superfície interpolada e variância da predição dos atributos de solos, da camada 15 a 30 cm de profundidade: capacidade de troca catiônica a pH 7 (CTC); matéria orgânica do solo (MOS); teor de areia (Areia); teor de silte (Silte) e teor de argila (Argila).

Tabela 2 - Estimativas dos parâmetros dos modelos dos semivariogramas ajustados e densidade de amostras requeridas para a predição espacial dos atributos de solos, da camada 15 a 30 cm de profundidade, da bacia do Rio Pongal.

Variável <sup>(1)</sup>	Modelo <sup>(2)</sup>	C0	C1	C0+C1	Alcance (m)	DE	Densidade (pontos ha <sup>-1</sup> )
pH	Esf	0,21	0,05	0,26	913,3	0,19	-
CTC	Exp	0,00	6,93	6,93	547,2	1,00	0,03
MOS	Exp	0,09	0,70	0,79	565,5	0,89	0,03
Areia	Exp	0,00	209,10	209,10	297,6	1,00	0,11
Silte	Exp	0,00	14,74	14,74	360,6	1,00	0,08
Argila	Exp	0,00	181,64	181,64	433,7	1,00	0,05

<sup>(1)</sup> pH: pH em água (-log[H<sup>+</sup>]); CTC: capacidade de troca catiônica a pH 7,0 (cmol<sub>c</sub> dm<sup>-3</sup>); MOS: matéria orgânica do solo (dag kg<sup>-1</sup>); Areia: teor de areia (dag kg<sup>-1</sup>); Silte: teor de silte (dag kg<sup>-1</sup>); Argila: teor de argila (dag kg<sup>-1</sup>). <sup>(2)</sup> Esf: esférico; Exp: exponencial. C0: Efeito pepita; C1: Contribuição; C0+C1: Patamar; DE: Dependência espacial (fraca = 0 a 0,25; moderada = 0,25 a 0,75; forte = 0,75 a 1,00).

#### 1.4 DISCUSSÃO

Os valores discrepantes (*outliers*) das variáveis CTC (2 pontos) e MOS (5 pontos) são relacionados às amostras coletadas em Gleissolos, localizados nas posições mais baixas da bacia hidrográfica. O processo de hidromorfismo a que estão sujeitos estes solos justificam os acúmulos de material orgânico com a elevação da CTC. O ponto que foi considerado *outlier* para a variável silte (maior teor de silte) foi também considerado como *outlier* para a variável argila (menor teor de argila) e foi coletado no Gleissolos. A variável areia apresentou um ponto discrepante localizado em uma pequena mancha de Espodossolo, formada em condição de drenagem deficiente, com a concentração de quartzo e a translocação de argilo-minerais e da matéria orgânica do solo (MOS).

Para adequada espacialização dos atributos de solos, as regiões que apresentam uma maior variabilidade de valores, no caso as áreas de Gleissolos, devem receber um maior adensamento de amostras. Nestes casos, o mapa de variância da predição, gerado no processo de krigagem, deve ser considerado no planejamento do adensamento das amostras.

A variável pH apresentou fraca dependência espacial. Neste caso, mais de 80% da variabilidade dos dados não é explicada pelo modelo teórico ajustado ao semivariograma, e não é recomendado o uso da geoestatística para a predição espacial, uma vez que o pré-requisito dos métodos geoestatísticos de krigagem é a existência de dependência espacial. Este resultado era esperado, uma vez que o pH é o atributo de solo mais sensível às ações humanas (antrópicas) na paisagem. O manejo das áreas agrícolas e pastagens, com a aplicação ou não de corretivos agrícolas, e a própria variabilidade natural, resultam em grandes variações nos valores do pH do solo.

As demais variáveis apresentaram forte dependência espacial ( $DE > 0,75$ ). As variáveis CTC e MOS apresentaram os maiores alcances, seguidas pela variável argila, silte e areia, que apresentou o menor alcance. No caso do silte e da areia, o espaçamento mínimo entre os pontos amostrais mostrou-se insuficiente para captar a variabilidade das variáveis em distâncias menores que o alcance. Assim, é necessária uma amostragem que permita um maior pareamento de pontos em distâncias inferiores ao alcance, para melhorar a modelagem e avaliação do semivariograma.

Com o objetivo de obter um esquema amostral ótimo para um conjunto de atributos do solo, Pereira *et al.* (2013) utilizaram a modelagem de semivariogramas escalonados para estabelecer os parâmetros **C0** (pepita), **C1** (contribuição) e **a** (alcance), em uma área de 42 hectares. Os alcances dos atributos químicos e físicos foram de 557 m e 716 m, respectivamente, com dependência espacial (DE) moderada em ambos os casos. Comparando os resultados, as variáveis CTC e MOS apresentaram alcances próximos do valor encontrado para os atributos químicos. No entanto, as variáveis areia, silte e argila apresentaram alcances menores que o encontrado para os atributos físicos. Neste caso, a área estudada por Pereira *et al.* (2013) possui tamanho reduzido e apresenta apenas um tipo de solo (Latosolo Vermelho Distrófico), que favoreceu a obtenção do maior valor para o alcance para os atributos físicos.

A densidade amostral, estimada a partir dos valores de alcance, apresentou valores entre 0,03 e 0,11 pontos por hectares. Esses valores são compatíveis com as densidades de observações requeridas em levantamentos semidetalhados de solos, que variam de 0,02 a 0,20 observações por hectares, segundo IBGE (2007).

O plano de amostragem, que utilizou grade regular hexagonal com pontos adicionais localizados em distâncias intermediárias ao espaçamento da grade, mostrou-se adequado para a avaliação de cinco das seis variáveis estudadas. O uso de grades regulares apresenta grande vantagem, uma vez que otimiza o número de pareamentos utilizados na determinação do semivariograma. Contudo, as grades regulares podem impor maiores dificuldades ao processo de coleta de amostras em campo, quando as áreas em estudo apresentam grande heterogeneidade topográfica, e de uso e cobertura das terras.

Apesar dos bons resultados obtidos na modelagem dos semivariogramas, as superfícies interpoladas para os atributos de solos (CTC, MOS, areia, silte e argila) apresentam elevada variância das predições (Figura 5). Os valores elevados de variância da predição são ocasionados, principalmente, pela densidade amostral usada ser menor

que a densidade amostral requerida para a predição. O adequado adensamento das amostras resulta na redução da variância das predições.

É importante relatar a existência de poucos estudos relacionados à variabilidade espacial dos atributos de solos em escalas intermediárias (regionais) de levantamento. A maioria dos estudos publicados foram realizados em pequenas áreas (escalas locais) que dificulta a comparação dos resultados aqui obtidos.

## 1.5 CONCLUSÕES

O pH é o atributo que apresenta a maior variabilidade espacial, entre os atributos de solos avaliados na área de estudo, exigindo uma densidade amostral muito alta para a predição espacial em escalas semidetalhadas, usando a krigagem ordinária.

A variabilidade espacial dos atributos CTC, MOS, areia, silte e argila, apresentou forte dependência espacial, que permite o uso de métodos geoestatísticos para predições não viesadas e com variância mínima, em escalas semidetalhadas de mapeamento.

O uso de técnicas geoestatísticas é viável para obtenção de informações espaciais sobre a variabilidade dos atributos de solos em áreas extensas. Os custos com a amostragem e as análises laboratoriais, exigidas para a aplicação da geoestatística, são compatíveis com os custos do mapeamento convencional (clássico), que requer densidade amostral equivalente, segundo IBGE (2007). Contudo, a geoestatísticas apresenta algumas vantagens sobre os métodos convencionais, tais como: reprodutibilidade das predições, maior detalhamento das informações geradas e estimativas das incertezas associadas às predições.

## REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO JUNIOR, W.; LAGACHERIE, P.; CHAGAS, C. D. S.; CALDERANO FILHO, B.; BHERING, S. B. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. **Geoderma**, v. 232-234, p. 479-486, 2014.

CPRM. **Programa Levantamentos Geológicos Básicos do Brasil**. Píuma. Folha SF.24-V-A-VI. Estado do Espírito Santo. Escala 1:100.000. Brasília, DF: DNPM/CPRM: 140 p. 1993.

EMBRAPA, Ed. **Manual de métodos de análise de solo**. Rio de Janeiro, RJ: Embrapa, p. 230, 2. ed. rev. ed. 2011.

ESRI. **ArcGIS Desktop for 10.2**. Environmental Systems Research Institute, Disponível em: < <http://www.esri.com> > Acesso em: junho de 2013.

GLOBALSOILMAP.NET. **Specifications Version 1 GlobalSoilMap.net products**. Disponível em: < <http://globalsoilmap.net/> > Acesso em: julho de 2013.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York, NY: Oxford University Press, 1997. 483 p.

IBGE. **Manual técnico de pedologia**. Rio de Janeiro, RJ: IBGE: 316 p. 2007.

PEREIRA, G. T.; SOUZA, Z. M.; TEIXEIRA, D. B.; MONTANARI, R.; MARQUES JR, J. Optimization of the sampling scheme for maps of physical and chemical properties estimated by kriging. **Revista Brasileira de Ciência do Solo**, v. 37, p. 1128-1135, 2013.

R DEVELOPMENT CORE TEAM. **The R Project for Statistical Computing**. Disponível em: < <http://www.r-project.org/> > Acesso em: novembro de 2013.

RADAMBRASIL. **Levantamento de Recursos Naturais Volume 32**. Folhas SF 23/24 Rio de Janeiro/Vitória: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Rio de Janeiro, RJ: Projeto RADAMBRASIL: 780 p. 1983.

RUIZ, H. A. Incremento da exatidão da análise granulométrica do solo por meio da coleta da suspensão. **Revista Brasileira de Ciência do Solo**, v. 29, p. 297-300, 2005.

SANTOS, R. D.; LEMOS, R. C.; SANTOS, H. G.; KER, J. C.; ANJOS, L. H. C.; SHIMIZU, S. H. **Manual de descrição e coleta de solo no campo**. 6. ed. rev. ampl. ed. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2013. 100 p.

VIEIRA, S. R. Geoestatística em estudos da variabilidade espacial do solo In: NOVAIS, R. F.; ALVARES V., V. H., *et al.* (Eds.). **Tópicos em Ciência do Solo**. 1 ed. Viçosa, MG: SBCS, v. 1, 2000. p. 1-54.

WEBSTER, R.; OLIVER, M. A. Sample Adequately to Estimate Variograms of Soil Properties. **Journal of Soil Science**, v. 43, n. 1, p. 177-192, 1992.

YFANTIS, E. A.; FLATMAN, G. T.; BEHAR, J. V. Efficiency of Kriging Estimation for Square, Triangular, and Hexagonal Grids. **Mathematical Geology**, v. 19, p. 183–205, 1987.

ZIMBACK, C. R. L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade do solo**. 2001. 114 p. Tese (Livre-Docência) - Faculdade de Ciências Agronômicas, Universidade Estadual Paulista, Botucatu, SP, 2001.

## **CAPÍTULO 2: SELEÇÃO DE COVARIÁVEIS AMBIENTAIS E PREDIÇÃO ESPACIAL DE ATRIBUTOS DE SOLOS USANDO REGRESSÃO LINEAR MÚLTIPLA E *RANDOM FOREST***

### **RESUMO**

A maioria dos métodos de mapeamento digital de atributos de solos relacionam diretamente covariáveis ambientais, obtidas de diferentes fontes, e atributos de solos. O uso de covariáveis ambientais pelos modelos preditivos visa à melhoria da exatidão das predições. As covariáveis mais promissoras para a espacialização de atributos de solos são aquelas geradas a partir da análise digital de terreno, de sensoriamento remoto do espectro visível ao infravermelho próximo (400-2500 nm) e de espectrometria de raios gama. Surge então a pergunta: qual covariável ou conjunto de covariáveis é melhor, em termos de usabilidade e exatidão, em matéria de mapeamento digital de solos? Os principais objetivos deste estudo foram: (1) identificar, entre quatro MDEs disponíveis, qual é mais adequado para a predição de atributos de solos (pH, CTC, MOS, areia, silte e argila); (2) selecionar o conjunto de covariáveis mais relevante para predição dos atributos de solos na Bacia do Rio Pongal; e (3) comparar o desempenho da Regressão Linear Múltipla (RLM) e *Random Forest* (RF) para a modelagem das superfícies dos atributos de solos. Os resultados apontam que as covariáveis derivadas do MDE 5 m (maior escala e resolução) são mais relevantes na predição dos atributos de solos estudados; a combinação de covariáveis ambientais de diferentes fontes (MDE, imagens de satélite, dados gamaespectrométricos) melhora os resultados dos modelos preditivos, com o uso de técnicas de mineração de dados; e que os modelos RF apresentam melhores resultados que os modelos RLM, na predição espacial de atributos de solos.

Termos para indexação: mapeamento digital de solos, mineração de dados, correlação ambiental, colinearidade.

### **SELECTION OF ENVIRONMENTAL COVARIATES AND SPATIAL PREDICTION OF SOILS PROPERTIES USING MULTIPLE LINEAR REGRESSION AND RANDOM FOREST**

#### **ABSTRACT**

Most methods of digital mapping of soils properties relate directly environmental covariates, obtained from different sources, and soil properties. The use of

environmental covariates by predictive models aims to improve the accuracy of predictions. The most promising covariates for spatialization of soil properties are those generated from digital terrain analysis, remote sensing from visible to near-infrared spectrum (400-2500 nm) and gamma spectrometry. The question then arises: which covariate or set of covariates is better in terms of usability and precision, in terms of digital soil mapping? The main objectives of this study were: (1) identify between four DEMs available which is more suitable for predicting soil properties (pH, CEC, SOM, sand, silt e clay); (2) select the set of covariates is more relevant to the prediction of soil properties in Pongal River Basin; and (3) compare the performance of Multiple Linear Regression (MLR) and Random Forest (RF) for modeling the surfaces of soil properties. The results indicate that the covariates derived from DEM 5 m (higher resolution) are more relevant in the prediction of the soil properties studied; the combination of environmental covariates of different sources (DEM, satellite images, gamma-ray spectrometric data) improves the results of predictive models, using data mining techniques; and the RF models show better results than the MLR models in the spatial prediction of soil properties.

Index terms: digital soil mapping, data mining, environmental correlation, collinearity.

## 2.1 INTRODUÇÃO

A maioria dos métodos de mapeamento digital de atributos de solos relacionam diretamente covariáveis ambientais, obtidos de diferentes fontes, e atributos de solos (Mendonça-Santos *et al.*, 2007). Especialmente em ambientes de escassez de dados de solos, procedimentos geoestatísticos e de regressão podem usar covariáveis ambientais para a predição espacial dos atributos de solos.

Os avanços tecnológicos do sensoriamento remoto propiciam um grande número de covariáveis ambientais, especialmente densas, e de fácil obtenção, que podem auxiliar a inferência, baseado em modelos estatísticos que produzem expressões quantitativas de variabilidade de solo e do seu erro (Rossel e Chen, 2011).

O uso de covariáveis ambientais pelos modelos preditivos visa à melhoria da exatidão das predições. As covariáveis mais promissoras para a espacialização dos atributos de solos são aquelas geradas a partir da análise digital de terreno, que refletem os processos geomorfológicos, climáticos e hidrológicos, de sensoriamento remoto do espectro visível ao infravermelho próximo (400-2500 nm), que fornecem informações

sobre a cor, óxidos de ferro, minerais de argila, carbonados e teor de matéria orgânica (Rossel e Behrens, 2010), e de espectrometria de raios gama, que detectam diferenças no tamanho de partículas e na mineralogia do material de origem (Wilford *et al.*, 1997). O uso dessas tecnologias exige registro cuidadoso entre observações de campo e outras coberturas digitais, que é obtido através do uso de sistemas de posicionamento global com correção diferencial (McKenzie e Ryan, 1999).

Na análise digital de terreno, diferentes resoluções do Modelo Digital de Elevação (MDE) produzem distintas informações sobre as variáveis de terreno (atributos do terreno). E embora a escolha da resolução espacial dependa da complexidade da paisagem e das variáveis de terreno utilizadas na modelagem, muitos trabalhos que utilizam dados em formato matricial (*raster*) optam por uma determinada resolução espacial, sem qualquer justificativa científica em favor da opção adotada (Ten Caten *et al.*, 2012).

Surge então a pergunta: qual covariável ou conjunto de covariáveis é melhor, em termos de usabilidade e exatidão, em matéria de mapeamento digital de solos? Técnicas de mineração de dados podem ajudar a responder esta questão. Para Minasny e McBratney (2007) mais importante que utilizar métodos estatísticos sofisticados para predição espacial dos atributos de solos, é o uso de covariáveis de elevada qualidade e maior grau de correlação com a variável primária. O uso de covariáveis de baixa qualidade não ajuda a melhorar a precisão e aumenta a complexidade dos modelos preditivos, podendo piorar os resultados da predição.

Assim, os principais objetivos deste estudo foram: (1) identificar, entre os quatro MDEs disponíveis, qual é mais adequado para a predição de atributos de solos; (2) selecionar o conjunto de covariáveis mais relevante para predição dos atributos de solos da Bacia do Rio Pongal; e (3) comparar o desempenho das técnicas de regressão linear múltipla e *random forest* para a modelagem dos atributos de solos.

## **2.2 MATERIAL E MÉTODOS**

### **2.2.1 Área de estudo**

A área de estudo compreende a Bacia do Rio Pongal, no município de Anchieta, estado do Espírito Santo, com aproximadamente 140 km<sup>2</sup>. Localizada na região de Mata Atlântica com clima tropical quente úmido. Atualmente, a área é predominante ocupada por pastagens, com pequenas lavouras de café e remanescentes florestais (Figura 1).

A complexidade pedológica da região é resultante da geologia diversificada somada a variedade de geformas do relevo.

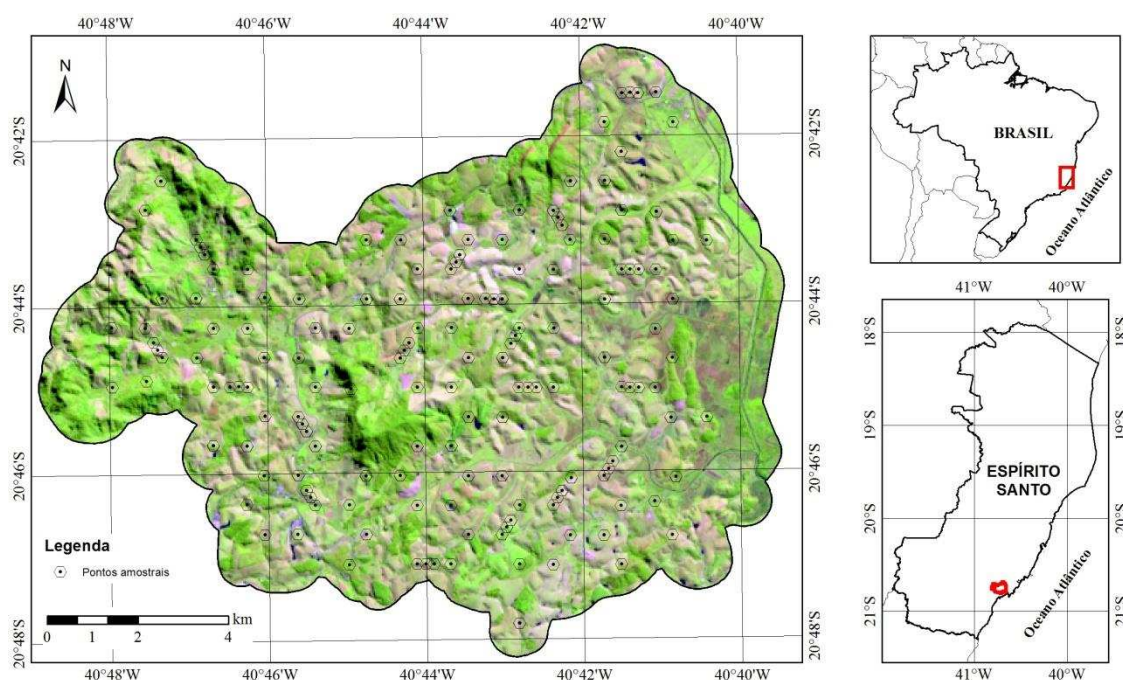


Figura 1 - Área de estudo e localização dos pontos de amostragem sobre imagem Landsat 8 composição falsa cor (bandas 6, 5 e 4).

### 2.2.2 Base de dados dos atributos de solos

A base de dados dos atributos de solos foi construída a partir de uma malha aproximadamente hexagonal, com 154 pontos amostrais, georreferenciados com receptor GPS no modo pós-processado com correção diferencial (DGPS). As amostras foram coletadas na camada entre 15 e 30 cm de profundidade (GlobalSoilMap.net, 2011).

Análises laboratoriais químicas e físicas de rotina foram realizadas, segundo métodos descritos por Ruiz (2005) e Embrapa (2011), para determinar os valores do pH em água, potássio (K), cálcio ( $\text{Ca}^{2+}$ ), magnésio ( $\text{Mg}^{2+}$ ), acidez potencial (H+Al), carbono orgânico e os teores de areia, silte e argila. Foram definidas como variáveis os atributos primários de solos: pH em água; capacidade de troca catiônica a pH 7 (CTC); matéria orgânica do solo (MOS); teor de areia, silte e argila; cujos resultados obtidos foram tabulados com suas respectivas coordenadas no sistema de projeção UTM fuso 24 sul, no datum SIRGAS 2000.

Os pontos com valores discrepantes (*outliers*) de cada variável foram removidos do conjunto de dados, utilizando o critério Z-Scores maior que três, mantendo apenas os valores compreendidos entre os limites da média mais ou menos três vezes o desvio padrão.

Para realizar os processos de predição e de avaliação dos modelos preditivos, os pontos amostrais foram divididos em dois conjuntos: um conjunto de pontos de validação, contendo 30 pontos, para ser usado no processo de avaliação da exatidão dos modelos preditivos. Os pontos restantes compõem o segundo conjunto de pontos (pontos de treinamento) que foram utilizados no processo de predição.

### 2.2.3 Covariáveis ambientais

Foram reunidas covariáveis ambientais com potencial para predição dos atributos de solos (Tabela 1). As covariáveis selecionadas foram aquelas comumente correlacionadas com os atributos de solo, de acordo com a literatura (McKenzie e Ryan, 1999; Henderson *et al.*, 2005; Pracilio *et al.*, 2006; Carvalho Junior *et al.*, 2014), e disponíveis para a área de estudo.

O primeiro conjunto de covariáveis foi gerado com base nos Modelos Digitais de Elevação (MDE) de diferentes fontes e resoluções espaciais: a) SRTM Digital Elevation Data v4.1 - 90 m (MDE 90 m); b) Aster GDEM v2 - 30 m (MDE 30 m); c) Base cartográfica IBGE na escala 1:50.000 interpolada - 10 m (MDE 10 m); d) Levantamento aerofotogramétrico na escala 1:10.000 - 5 m (MDE 5 m). Todos os modelos foram hidrológicamente corrigidos, identificando e preenchendo os sumidouros (*sinks*).

Quatorze atributos de terreno foram obtidos a partir de cada MDE: 1) elevação, 2) inclinação (*slope*), 3) face de exposição (*aspecto*), 4) curvatura, 5) perfil de curvatura, 6) multirresolução de topos aplainados (MRRTF - *Multi-Resolution Ridge Top Flatness*), 7) multirresolução de fundos de vales aplainados (MRVBF - *Multi-Resolution Valley Bottom Flatness*), 8) posição de declividade média (MSP - *Mean Position of Declivity*), 9) índice de rugosidade do terreno (TRI - *Terrain Ruggedness Index*), 10) índice topográfico de umidade (TWI - *Topographic Wetness Index*), 11) radiação solar total, 12) radiação solar direta, 13) radiação solar difusa e 14) duração da radiação solar direta. Os dez primeiros atributos topográficos foram gerados no *System for Automated Geoscientific Analyses* (SAGA DEVELOPMENT TEAM, 2014) e os quatro últimos foram gerados no programa Arcgis 10.2, extensão *Spatial Analyst Tools* (ESRI, 2013).

O segundo conjunto, com seis covariáveis, foi obtido a partir dos dados gamaespectrométricos do contorno radiométrico, fornecidos pelo CPRM (Serviço Geológico do Brasil): 1) concentração do elemento tório (eTh), 2) concentração do elemento urânio (eU), 3) concentração do elemento potássio (Kperc), 4) razão eTh/K,

5) razão eU/K e 6) razões eU/eTh. As seis covariáveis são dados do Projeto Aerogeofísico do Espírito Santo, com resolução espacial de 125m (CPRM, 2010). A radiação gama (comprimentos de onda em torno de 10-3nm), emitida por rochas e solo, é proveniente da radioatividade natural do decaimento dos elementos potássio (K), urânio (U) e tório (Th) (Wilford *et al.*, 1997).

Tabela 1 - Covariáveis ambientais avaliadas nas predições.

Covariável	Relação
Elevação	Clima, vegetação, potencial energético.
Face de exposição (Aspecto)	Radiação solar, evapotranspiração, distribuição e abundância de flora e fauna.
Declividade ( <i>Slope</i> )	Fluxos de superfície e subsuperfície, velocidade de fluxo e a taxa de erosão, precipitações, vegetação, geomorfologia, conteúdo de água no solo, capacidade de uso da terra.
Curvatura e perfil de curvatura	Fluxo convergente/divergente, conteúdo de água no solo, características do solo, aceleração do fluxo, taxa de erosão/deposição e geomorfologia.
Multirresolução de topos aplainados (MRRTF)	Partes superiores da paisagem e áreas degradadas.
Multirresolução de fundos de vales aplainados (MRVBF)	Áreas de deposição e erosão.
Posição de declividade média (MSP)	Posição de altitude e inclinação.
Índice de rugosidade do terreno (TRI)	Rugosidade e heterogeneidade topográfica.
Índice topográfico de umidade (TWI)	Umidade do solo.
Radiação solar total, direta e difusa, duração da radiação direta	Evapotranspiração, distribuição e abundância de flora e fauna e intemperismo.
Bandas 2 a 7 Landsat 8, índice de vegetação por diferença normalizada (NDVI), índices minerais de argila (CMI), índice de óxido de ferro (IOI), B4/B3, B4/B7	Uso e cobertura, tipo de vegetação, ciclagem de nutrientes, umidade do solo, minerais de argila e óxido de ferro.
Concentração do elemento tório (eTh)	Mineralogia mais resistente ao intemperismo e complexos orgânicos.
Concentração do elemento urânio (eU)	Mineralogia mais resistente ao intemperismo e complexos orgânicos.
Concentração do elemento potássio (Kperc)	Material de origem (rochas ácidas) e grau de intemperismo.

Adaptada e ampliada de Carvalho Junior *et al.* (2014).

Dados do sensor OLI do satélite Landsat-8, de junho de 2014, foram utilizados no terceiro conjunto com 11 covariáveis: 1) banda 2 (azul, 0,45-0,51  $\mu\text{m}$ ), 2) banda 3

(verde, 0,53-0,59  $\mu\text{m}$ ), 3) banda 4 (vermelho, 0,64-0,67  $\mu\text{m}$ ), 4) banda 5 (NIR, 0,85-0,88  $\mu\text{m}$ ), 5) banda 6 (SWIR 1, 1,57-1,65  $\mu\text{m}$ ), 6) banda 7 (SWIR 2, 2,11-2,29  $\mu\text{m}$ ), 7) índices minerais de argila (*clay mineral index*, CMI), 8) índice de óxido de ferro (*iron oxide index*, IOI); 9) índice de vegetação por diferença normalizada (*normalized difference vegetation index*, NDVI), 10) razão entre a banda 4 e a banda 3 e 11) razão entre a banda 4 e a banda 7 (Cunha, 2013; Carvalho Junior *et al.*, 2014).

#### **2.2.4 Correlação entre atributos de solo e covariáveis ambientais**

Para identificar quais covariáveis ambientais estão correlacionadas com os atributos de solo foi aplicado o teste clássico de correlação entre dados pareados, com base no Coeficiente de Correlação Linear de Pearson ( $r$ ) e no Teste  $t$ . O resultado do teste é a probabilidade da hipótese de não correlação entre um determinado atributo de solo e a covariável ambiental (Ciampalini *et al.*, 2012).

O software estatístico R (R DEVELOPMENT CORE TEAM, 2013) foi usado para calcular o Coeficiente de Correlação Linear de Pearson ( $r$ ) e o valor de significância do Teste T ( $p$ -value), com a função `cor.test` do pacote `stats` (*The R Stats Package*).

Este é um passo importante para identificação das covariáveis mais adequadas ao modelo de predição, bem como para comparação dos resultados obtidos ao fim do processo.

#### **2.2.5 Seleção do MDE**

Para a predição dos atributos de solo, foi selecionado o Modelo Digital de Elevação (MDE), entre os quatro MDEs disponíveis, que apresentou o maior número de atributos de terreno correlacionados com os atributos de solo (pH, CTC, MOS, areia, silte e argila), com base no Coeficiente de Correlação de Linear de Pearson. Foi quantificado o número de atributos de terreno por MDE com coeficiente de correlação significativo a 0,1% ( $p$ -value < 0,001).

#### **2.2.6 Modelos preditivos**

##### **2.2.6.1 Regressão Linear Múltipla**

O modelo de Regressão Linear Múltipla (RLM) pode ser escrito na forma matricial:

$$V^* = \alpha + U\beta + \epsilon \quad (1)$$

onde  $\alpha$  e  $\beta$  são vetores dos coeficientes determinados pelo método de mínimos quadrados,  $U$  é a matriz com os valores da covariável, e  $\epsilon$  é o vetor de erro da predição, que tem média zero e variância constante.

A partir do conjunto com 31 covariáveis (14 derivadas do MDE selecionado, 6 derivadas dos dados gamaespectrometricos e 11 derivadas das imagens Landsat 8) foram selecionadas as covariáveis predictoras para os modelos de RLM. Para simplificação dos modelos de RLM, foi combinado o método *Stepwise*, que usa o AIC (*Akaike Information Criterion*) na escolha do modelo, com o fator de inflação da variância (VIF - *Variance Inflation Factor*) para o diagnóstico da colinearidade (Fox e Monette, 1992) e a medida da importância relativa para modelos lineares (Lindeman *et al.*, 1980), para remover as covariáveis colineares de menor importância, conforme diagrama (Figura 2).

Para realizar a análise e a seleção de covariáveis foi construído um *script* no software estatístico R (R DEVELOPMENT CORE TEAM, 2013), usando funções dos pacotes *stats* (*The R Stats Package*), *relaimpo* (*Relative importance of regressors in linear models*) e *car* (*Companion to Applied Regression*).

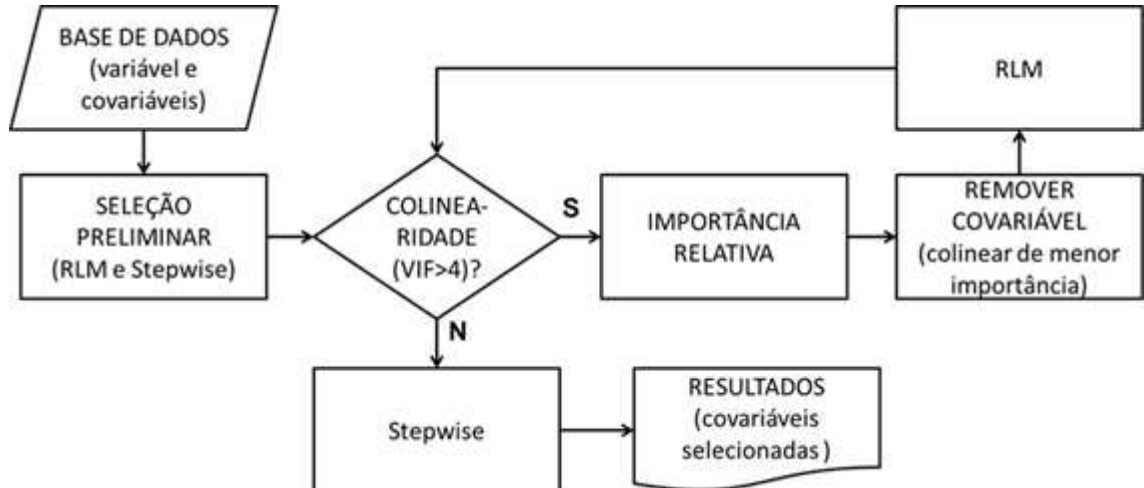


Figura 2 - Diagrama do processo de seleção SRC (*Stepwise Removendo a Colinearidade*) de covariáveis ambientais para a predição de atributos de solos.

### 2.2.6.2 *Random Forest*

A *Random Forest* (RF) é uma técnica de inteligência artificial, baseada em árvores de regressão, usada para a análise multivariada de dados não paramétricos e não lineares, que incorpora os conceitos de mineração de dados (Grimm *et al.*, 2008; Ließ *et al.*, 2012).

O modelo RF combina um conjunto de árvores de regressão a fim de melhorar a exatidão da predição. Cada árvore é construída a partir da seleção aleatória de covariáveis preditoras e pontos de treinamento. O algoritmo RF é bastante robusto, pouco sensível a ruídos nos dados e ao *overfitting* (super ajuste). Outra característica importante é a capacidade de fornecer medidas de importância das covariáveis na predição (Heung *et al.*, 2014).

A RF fornece uma estimativa confiável de erro, usando OOB (*out-of-bag*), em que um terço dos dados de entrada é excluído da predição para serem usados na estimativa do erro de predição, sem a necessidade de um conjunto de dados extra (dados de validação) para estimar o erro da predição. Esta vantagem é particularmente interessante para a ciência de solos, uma vez que a coleta de amostras e as análises laboratoriais são caras e demoradas (Grimm *et al.*, 2008).

Três parâmetros são definidos pelos usuários no processamento: o número de árvores da floresta (*ntree*), o número mínimo de nós terminais (*nodesize*) e número de covariáveis selecionadas aleatoriamente em divisão (*mtry*). Foi definido *nodesize* igual a cinco e *mtry* igual a um terço do número total de preditores, que são valores padrões para os problemas de árvores de regressão. Para o *mtry* foi adotado o valor de 1000 árvores (Grimm *et al.*, 2008).

Os modelos *Random Forest* para a predição dos atributos de solo (pH, CTC, MOS, Areia, Silte e Argila) foram gerados a partir de 31 covariáveis preditoras, usadas também na Regressão Linear Múltipla (RLM).

Para verificar os resultados obtidos no processo de seleção de covariáveis foi usada a medida de importância da variável, calculada pela diferença do erro OOB do *Random Forest*. Os resultados da medida de importância do *Random Forest* são apresentados na forma de gráfico de ranking de covariáveis por incremento de erro quadrático médio (MSE - *Mean Squared Error*).

As análises foram realizadas com o pacote *randomForest* versão 4.6-10 no software estatístico R (R DEVELOPMENT CORE TEAM, 2013).

### **2.2.6.3 Avaliação da exatidão dos modelos preditivos**

O desempenho dos modelos preditivos foi avaliado por meio do coeficiente de determinação ( $R^2$ ), obtido na regressão linear entre valores preditos e observados do conjunto de pontos de validação.

A exatidão da predição dos modelos foi avaliada por meio da raiz quadrada do erro médio quadrático (RMSE - *Root Mean Squared Error*), calculada por:

$$RMSE = \sqrt{\frac{1}{n} \sum [z^* - z]^2} \quad (2)$$

em que  $z^*$  é o valor predito e  $z$  é o valor observado nos pontos de validação.

## 2.3 RESULTADOS

### 2.3.1 Estatística descritiva

A base de dados dos atributos de solos continha um *outlier* (valores discrepantes) para as variáveis pH, areia, silte e argila, dois *outliers* para variável CTC, e cinco *outliers* para a variável MOS. Os *outliers* de cada variável foram excluídos e o resumo dos dados de solos, após a remoção dos valores discrepantes, é apresentado na Tabela 2.

Tabela 2 - Estatística descritiva dos dados dos atributos de solos, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal.

Variável <sup>(1)</sup>	N	Média	s	CV	Mínimo	Máximo
pH (-log [H <sup>+</sup> ])	153	5,06	0,51	0,10	3,63	6,53
CTC (cmol <sub>c</sub> dm <sup>-3</sup> )	152	6,51	2,63	0,40	2,58	21,92
MOS (dag kg <sup>-1</sup> )	149	2,33	0,91	0,39	0,52	7,93
Areia (dag kg <sup>-1</sup> )	153	42,83	14,97	0,35	1,00	88,00
Silte (dag kg <sup>-1</sup> )	153	8,19	4,06	0,50	1,00	25,00
Argila (dag kg <sup>-1</sup> )	153	48,41	13,74	0,28	7,00	84,00

<sup>(1)</sup> pH: pH em água; CTC: capacidade de troca catiônica a pH 7,0; MOS: matéria orgânica do solo; Areia: teor de areia, Silte: teor de silte; Argila: teor de argila; n: número de observações da base de dados dos atributos de solos; s: desvio padrão; CV: coeficiente de variação.

### 2.3.2 Análise da correção linear

Os valores estimados para as correlações entre variáveis e covariáveis flutuaram no intervalo de -0,40 a 0,69 (Tabela 3). Valores positivos indicam correlação direta (o aumento de uma indica aumento da outra), enquanto negativos indicam correlação inversa (o aumento de uma indica diminuição da outra). Valores de  $r$  próximos de zero indicam que a variável não é correlacionada linearmente com a covariável.

Tabela 3 - Coeficiente de Correlação Linear de Pearson entre variáveis e covariáveis ambientais e significância dos coeficientes.

Covariáveis <sup>(1)</sup>	pH	CTC	MOS	Areia	Silte	Argila
----- MDE 90 m -----						
Elev90	-0,05	-0,17	-0,23	0,16	0,10	-0,18
Aspect90	0,00	-0,14	-0,21	0,15	-0,04	-0,09
Slope90	0,02	-0,21	<b>-0,29*</b>	0,14	0,00	-0,13
Curvat90	-0,04	-0,02	0,06	-0,01	-0,13	0,03
ProfCurv90	-0,05	0,00	0,10	0,00	-0,19	0,06
MRRTF90	0,14	0,03	0,04	0,07	-0,20	0,00
MRVBF90	-0,14	<b>0,42*</b>	<b>0,36*</b>	-0,22	0,24	0,12
MSP90	-0,09	0,23	0,06	-0,02	<b>0,33*</b>	-0,14
TRI90	0,04	-0,23	<b>-0,31*</b>	0,17	-0,03	-0,15
TWI90	-0,23	<b>0,46*</b>	<b>0,43*</b>	<b>-0,37*</b>	<b>0,31*</b>	0,24
SolarRad90	0,00	0,11	0,08	-0,01	-0,06	0,00
DiretRad90	0,00	0,11	0,06	0,01	-0,04	-0,03
DiffuRad90	0,02	0,05	0,20	-0,16	-0,13	0,19
DiretDur90	0,00	0,12	0,21	-0,21	-0,12	0,23
----- MDE 30 m -----						
Elev30	-0,04	-0,16	-0,21	0,13	0,12	-0,14
Aspect30	-0,08	-0,22	-0,11	<b>0,30*</b>	-0,15	-0,22
Slope30	-0,12	-0,25	-0,18	0,08	-0,03	-0,02
Curvat30	-0,02	0,05	0,04	0,00	-0,03	0,02
ProfCurv30	0,05	0,07	0,07	-0,03	-0,07	0,07
MRRTF30	0,25	0,21	0,07	-0,08	-0,07	0,12
MRVBF30	-0,17	<b>0,53*</b>	<b>0,34*</b>	<b>-0,30*</b>	<b>0,38*</b>	0,13
MSP30	0,04	0,18	0,09	0,10	0,05	-0,14
TRI30	-0,12	<b>-0,27*</b>	-0,19	0,07	0,00	-0,03
TWI30	-0,10	<b>0,53*</b>	<b>0,30*</b>	<b>-0,27*</b>	<b>0,33*</b>	0,14
SolarRad30	0,07	0,16	0,14	-0,02	-0,06	0,01
DiretRad30	0,06	0,14	0,11	0,00	-0,05	-0,01
DiffuRad30	0,11	0,20	0,26	-0,16	-0,12	0,18
DiretDur30	0,04	0,26	<b>0,30*</b>	-0,25	-0,05	0,26
----- MDE 10 m -----						
Elev10	-0,01	-0,16	-0,19	0,11	0,11	-0,11
Aspect10	-0,04	-0,12	-0,11	0,21	-0,08	-0,21
Slope10	-0,14	-0,19	-0,17	0,05	0,00	-0,01
Curvat10	0,02	-0,02	0,03	-0,06	-0,02	0,08
ProfCurv10	-0,02	0,05	0,12	-0,06	0,00	0,08
MRRTF10	0,02	0,15	0,02	-0,14	0,01	0,15
MRVBF10	-0,16	<b>0,35*</b>	0,22	-0,19	0,24	0,07
MSP10	0,06	0,12	0,01	0,01	0,15	-0,06
TRI10	-0,15	-0,18	-0,18	0,05	0,01	-0,02
TWI10	-0,16	<b>0,45*</b>	<b>0,30*</b>	-0,23	<b>0,33*</b>	0,08
SolarRad10	0,14	0,15	0,15	-0,05	-0,04	0,03
DiretRad10	0,13	0,14	0,12	-0,02	-0,02	0,01
DiffuRad10	0,20	0,17	0,25	-0,18	-0,11	0,20
DiretDur10	0,14	0,21	<b>0,27*</b>	-0,25	-0,05	0,26
----- MDE 5 m -----						
Elev05	-0,01	-0,18	-0,21	0,12	0,10	-0,12
Aspect05	0,15	<b>-0,34*</b>	-0,19	0,24	<b>-0,27*</b>	-0,14
Slope05	-0,13	<b>-0,39*</b>	<b>-0,38*</b>	0,13	0,01	-0,09
Curvat05	-0,05	0,00	0,03	-0,25	-0,05	<b>0,31*</b>
ProfCurv05	-0,03	0,08	0,08	<b>-0,30*</b>	-0,07	<b>0,36*</b>
MRRTF05	0,03	<b>0,40*</b>	0,25	-0,24	0,04	0,26
MRVBF05	-0,20	<b>0,64*</b>	<b>0,45*</b>	<b>-0,40*</b>	<b>0,46*</b>	0,20
MSP05	-0,03	<b>0,27*</b>	<b>0,27*</b>	-0,08	0,09	0,01
TRI05	-0,14	<b>-0,38*</b>	<b>-0,38*</b>	0,13	0,02	-0,09
TWI05	-0,18	<b>0,69*</b>	<b>0,56*</b>	<b>-0,38*</b>	<b>0,39*</b>	0,21
SolarRad05	0,10	0,18	0,20	-0,03	-0,13	0,05

Covariáveis <sup>(1)</sup>	pH	CTC	MOS	Areia	Silte	Argila
DiretRad05	0,09	0,15	0,16	0,00	-0,12	0,02
DiffuRad05	0,15	0,26	<b>0,36*</b>	-0,17	-0,16	0,21
DiretDur05	0,11	<b>0,27*</b>	<b>0,36*</b>	<b>-0,30*</b>	-0,08	<b>0,33*</b>
----- Gamaespectrometria -----						
eTh	0,01	-0,01	0,01	-0,01	-0,06	0,05
eU	-0,02	-0,07	-0,09	0,10	0,09	-0,09
Kperc	0,07	-0,04	-0,11	0,18	0,23	-0,23
ThKrazao	-0,03	-0,04	0,01	-0,11	-0,25	0,17
UKrazao	-0,14	-0,10	-0,02	-0,08	-0,25	0,16
UThrazao	-0,05	-0,04	-0,08	0,10	0,14	-0,12
----- Landsat 8 -----						
lsB2	0,16	0,06	0,15	-0,10	-0,04	0,13
lsB3	0,20	0,01	0,11	-0,05	-0,05	0,08
lsB4	0,12	-0,01	0,10	-0,09	0,01	0,11
lsB5	0,19	0,05	0,03	0,09	-0,10	-0,06
lsB6	0,13	0,00	0,10	-0,06	-0,04	0,09
lsB7	0,12	0,01	0,13	-0,08	-0,02	0,11
lsCMI	0,14	-0,02	0,05	0,05	-0,11	-0,01
lsIOI	0,10	-0,03	0,08	-0,08	0,02	0,10
lsNDVI	0,13	0,06	-0,03	0,16	-0,13	-0,15
lsrB4B3	-0,03	-0,02	0,08	-0,14	0,09	0,15
lsrB4B7	-0,12	-0,07	-0,17	0,05	0,06	-0,07

pH: pH em água ( $-\log[H^+]$ ); CTC: capacidade de troca catiônica a pH 7,0 ( $\text{cmol. dm}^{-3}$ ); MOS: matéria orgânica do solo ( $\text{dag kg}^{-1}$ ); Areia: teor de areia ( $\text{dag kg}^{-1}$ ), Silte: teor de silte ( $\text{dag kg}^{-1}$ ); Argila: teor de argila ( $\text{dag kg}^{-1}$ ); \* Coeficiente de correlação (r) significativo ao nível de 0,1% ( $p\text{-value} < 0,001$ ). <sup>(1)</sup> Elev: elevação; Aspect: face de exposição; Slope: inclinação; Curvat: Curvatura; ProfCurv: perfil de curvatura; MRRTF: *Multi-Resolution Ridge Top Flatness*; MRVBF: *Multi-Resolution Valley Bottom Flatness*; MSP: *Mean Position of Declivity*; TRI: *Terrain Ruggedness Index*; TWI: *Topographic Wetness Index*; SolarRad: radiação solar total; DiretRad: radiação solar direta; DiffuRad: radiação solar difusa; DiretDur: duração da radiação solar direta; eTh: concentração do elemento tório; eU: concentração do elemento urânio; Kperc: concentração do elemento potássio; ThKrazao: razão eTh/K; UKrazao: razão eU/K; UThrazao: razões eU/eTh; lsB2: bandas 2 Landsat 8; lsB3: bandas 3 Landsat 8; lsB4: bandas 4 Landsat 8; lsB5: bandas 5 Landsat 8; lsB6: bandas 6 Landsat 8; lsB7: bandas 7 Landsat 8; lsCMI: *clay mineral index*; lsIOI: *iron oxide index*; lsNDVI: *normalized difference vegetation index*; lsB4B3: razão banda 4/banda 3; lsB4B7: razão banda 4/banda 7.

O número de atributos de terreno de cada MDE que apresentou Coeficiente de Correlação de Linear de Pearson (r) significativo ao nível de 0,1% pelo Teste t é apresentado na Figura 3.

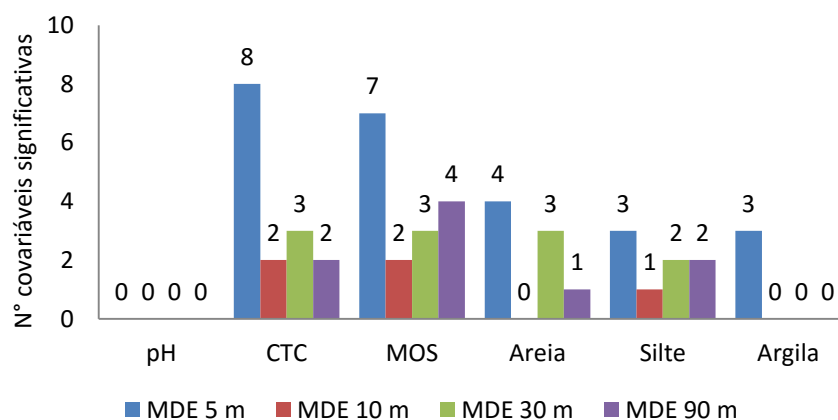


Figura 3 - Número de covariáveis ambientais por MDE significativamente correlacionadas às variáveis ( $p\text{-value} < 0,001$ ).

O MDE 5 m apresentou o maior número de atributos de terreno com coeficiente de correlação significativo em relação às variáveis, exceto para a variável pH, que não apresentou nenhuma covariável com correlação significativa ao nível de 0,1%. De maneira geral, os maiores valores absolutos do Coeficiente de Correlação Linear de Pearson entre as variáveis e covariáveis são observados no conjunto dos atributos de terreno derivado do MDE 5 m (Tabela 3). Por essas razões, os atributos de terreno derivados do MDE 5 m foram selecionados, entre os quatro MDEs disponíveis, para compor o conjunto de covariáveis preditoras para a Regressão Linear Múltipla (RLM) e para a *Random Forest* (RF).

### 2.3.3 Seleção de covariáveis ambientais para regressão linear múltipla

No processo de seleção de covariáveis ambientais para a RLM foram utilizadas as covariáveis do MDE 5 m (MDE selecionado) juntamente com as covariáveis derivadas de dados gamaespectrométricos e de dados do satélite Landsat 8, que somam 31 covariáveis.

Os resultados da seleção de covariáveis pelo método *Stepwise*, removendo a colinearidade (SRC), apresentam a combinação de covariáveis derivadas do MDE 5 m, com uma ou mais covariáveis derivadas de dados gamaespectrométricos e/ou Landsat 8 (Tabela 4).

Tabela 4 - Covariáveis selecionadas pelo modelo de regressão linear múltipla *Stepwise* removendo a colinearidade, gerados a partir de covariáveis ambientais, para a predição dos atributos de solos, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal.

Variável <sup>(1)</sup>	R <sup>2</sup> ajustado <sup>(2)</sup>	N <sup>o</sup> Covar. <sup>(3)</sup>	Covariáveis selecionadas <sup>(4)</sup>
pH (-log[H <sup>+</sup> ])	0,34	7	UKrazao + IsB3 + Curvat05 + MSP05 + TRI05 + TWI05 + SolarRad05
CTC (cmol <sub>c</sub> dm <sup>-3</sup> )	0,53	3	eU + ProfCurv05 + TWI05
MOS (dag kg <sup>-1</sup> )	0,34	2	IsrB4B3 + TWI05
Areia (dag kg <sup>-1</sup> )	0,31	4	IsNDVI + ProfCurv05 + MRVBF05 + DiretDur05
Silte (dag kg <sup>-1</sup> )	0,37	6	eU + IsIOI + Elev05 + Slope05 + MRRTF05 + MRVBF05
Argila (dag kg <sup>-1</sup> )	0,25	5	Slope05 + ProfCurv05 + MRRTF05 + MRVBF05 + DiretDur05

<sup>(1)</sup> pH: pH em água; CTC: capacidade de troca catiônica a pH 7,0; MOS: matéria orgânica do solo; Areia: teor de areia, Silte: teor de silte; Argila: teor de argila. <sup>(2)</sup> Coeficiente de determinação ajustado. <sup>(3)</sup> Número de covariáveis do modelo. <sup>(4)</sup> eU: concentração do elemento urânio; UKrazao: razão eU/K; IsB3: bandas 3 Landsat 8; IsB4B3: razão banda 4/banda 3 Landsat 8; IsIOI: *iron oxide index*; IsNDVI: *normalized difference vegetation index*; Elev05: elevação; Slope05: inclinação; Curvat05: Curvatura; ProfCurv05: perfil de curvatura; MRRTF05: Multi-Resolution Ridge Top Flatness; MRVBF05: Multi-Resolution Valley Bottom Flatness; MSP05: Mean Position of Declivity; TRI05: Terrain Ruggedness Index; TWI05: Topographic Wetness Index; SolarRad: radiação solar total; DiretDur05: duração da radiação solar direta.

### 2.3.4 Importância das covariáveis ambientais nos modelos *Random Forest*

O *ranking* com as 10 covariáveis mais importante para a predição dos atributos de solos estudados, segundo método *Random Forest*, é apresentado na Figura 4. Para todas as variáveis (pH, CTC, MOS, Areia, Silte e Argila) o *ranking* aponta que as covariáveis mais importantes são as derivadas do MDE 5 m, que ocupam, no mínimo, as quatro primeiras posições do *ranking* de cada variável.

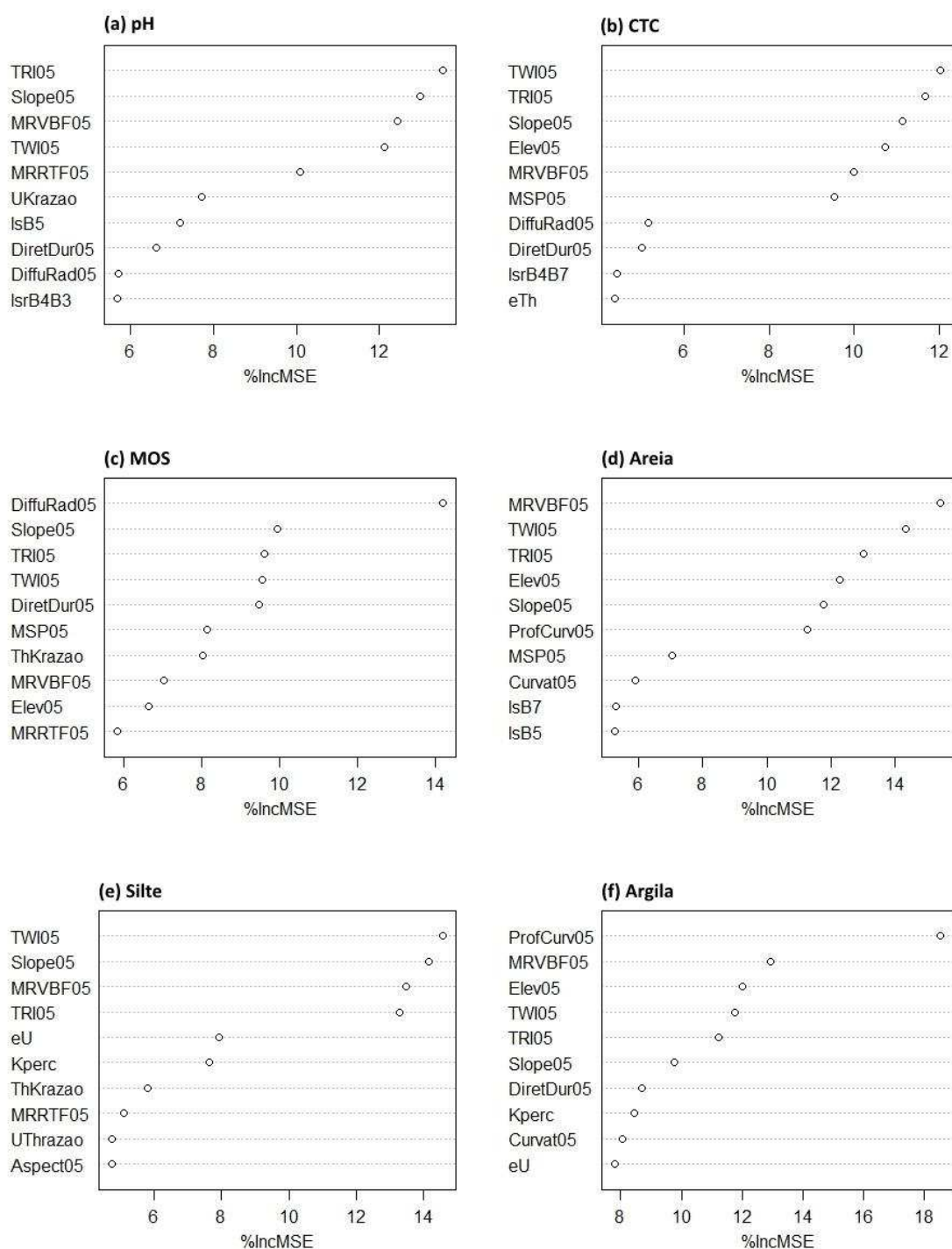


Figura 4 - *Ranking* da medida de importância das covariáveis ambientais para a predição dos atributos de solos pelo método *Random Forest*.

### 2.3.5 Mapa digital de atributos de solos e exatidão dos modelos preditivos

Para visualizar os resultados dos modelos RLM e *Random Forest* foram preditos os valores das variáveis para os pontos da grade de predição espacialmente distribuída por toda área de estudo (Figura 5 e 6).

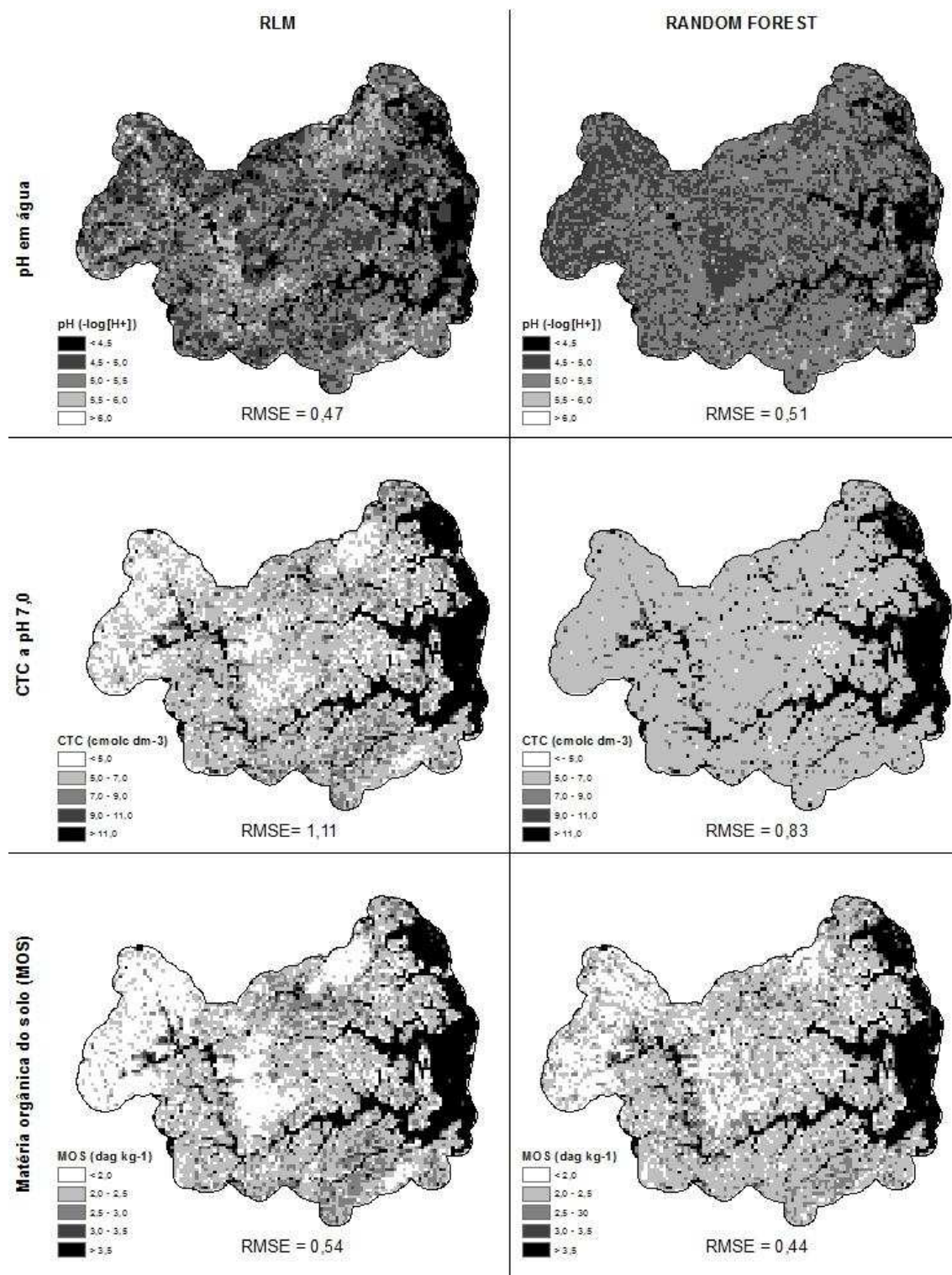


Figura 5 - Predição espacial de atributos químicas dos solos, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por Regressão Linear Múltipla (RLM) e *Random Forest*.

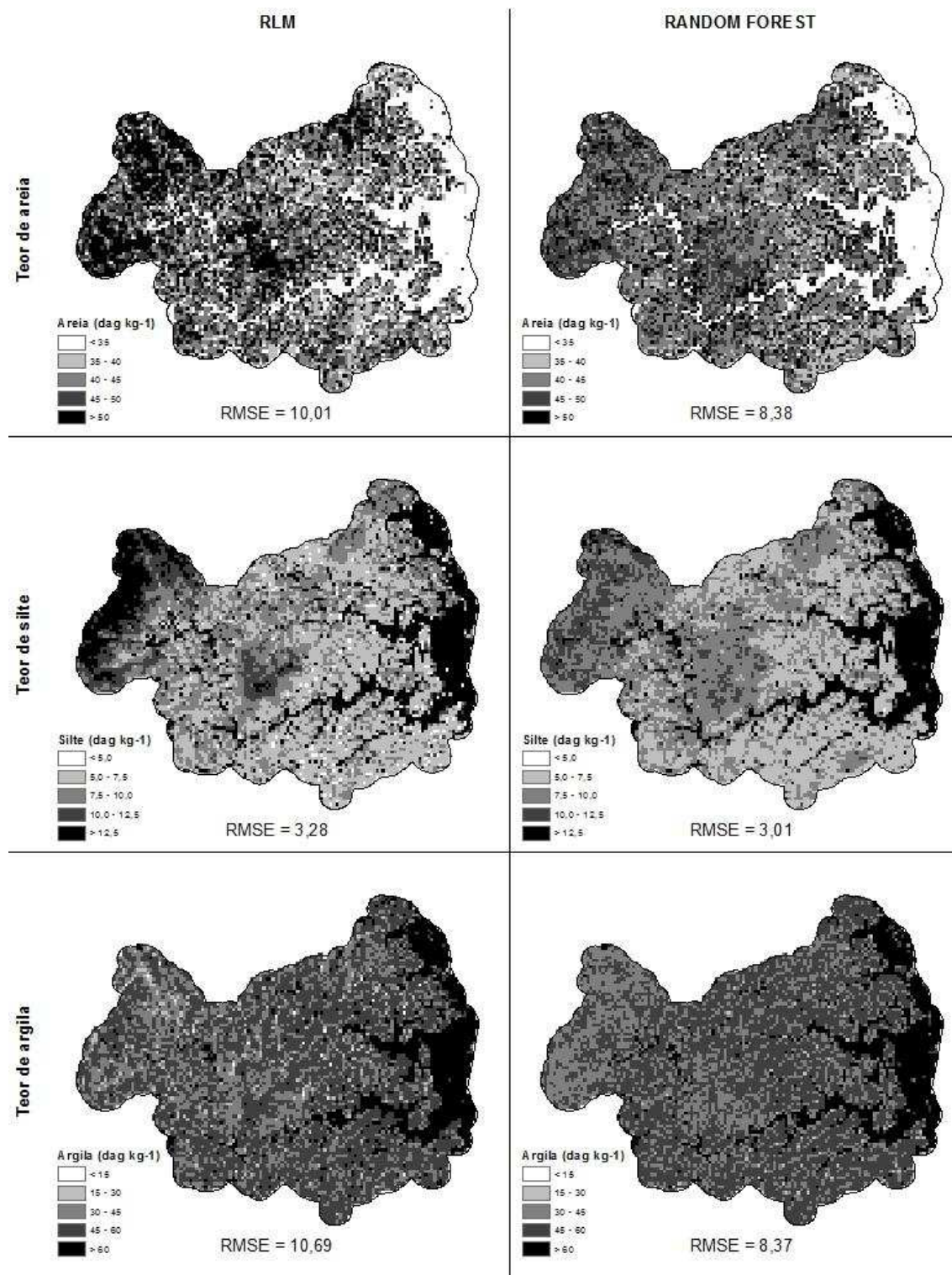


Figura 6 - Predição espacial de atributos físicos dos solos, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por Regressão Linear Múltipla (RLM) e *Random Forest*.

Com exceção da variável pH, os modelos preditivos da *Random Forest* apresentaram os melhores resultados, apresentando os maiores coeficientes de ajuste do modelo ( $R^2$ ) e menores RMSE, se comparado com os valores a RLM (Tabela 5). Os

resultados foram obtidos avaliando valores preditos e observados para o conjunto de pontos de validação.

Tabela 5 - Exatidão das predições dos atributos de solos, da camada de 15 a 30 cm de profundidade, da bacia do Rio Pongal.

Variável	Modelo	R <sup>2</sup>	RMSE
pH (-log[H <sup>+</sup> ])	RLM	0,16	0,47
	RF	0,00	0,51
CTC (cmol <sub>c</sub> dm <sup>-3</sup> )	RLM	0,42	1,11
	RF	0,54	0,83
MOS (dag kg <sup>-1</sup> )	RLM	0,10	0,54
	RF	0,30	0,44
Areia (dag kg <sup>-1</sup> )	RLM	0,10	10,32
	RF	0,22	8,38
Silte (dag kg <sup>-1</sup> )	RLM	0,03	3,28
	RF	0,14	3,01
Argila (dag kg <sup>-1</sup> )	RLM	0,19	9,73
	RF	0,34	8,37

pH: pH em água; CTC: capacidade de troca catiônica a pH 7; MOS: matéria orgânica do solo; Areia: teor de areia, Silte: teor de silte; Argila: teor de argila; RLM: Regressão Linear Múltipla; RF: *Random Forest*; R<sup>2</sup>: Coeficiente de determinação. RMSE: *Root Mean Squared Error*.

## 2.4 DISCUSSÃO

Os atributos de terreno, gerados a partir do MDE 5 m apresentaram, em geral, os maiores valores absolutos nas estimativas de correlação linear entre variáveis e covariáveis, em relação aos MDEs de menores resoluções espacial (MDE 10 m, MDE 30 m e MDE 90 m). O maior número de atributos de terreno com correlação significativa ao nível de 0,1% e os maiores valores absolutos dos coeficientes de correlação justificam a seleção do MDE 5 m para o processo de predição espacial dos atributos de solos.

Entre as covariáveis do MDE, merece especial destaque o TWI (índice topográfico de umidade) e o MRVBF (multi resolução fundo de vale aplainado), que apresentaram maior número de resultados significativos (p-value < 0,001) de correlação com as variáveis, nas diferentes resoluções do MDE. TWI e MRVBF também aparecem nas primeiras posições do *Ranking* da medida de importância das covariáveis ambientais do método *Random Forest* para a predição dos atributos de solos avaliados.

A covariável TWI é relacionada à umidade e ao acúmulo de água no solo, enquanto a MRVBF pode ser relacionada a fatores de pedogênese, identificando regiões de predomínio de erosão e regiões de deposição de material (McKenzie e Gallant, 2007). Analisando os resultados, observa-se correlação positiva das duas covariáveis com a CTC, MOS e teor de silte, e negativa com o teor de areia. Nas regiões favoráveis

ao acúmulo de umidade e água (hidromorfismo), predomina a sedimentação de materiais finos e o acúmulo de material orgânico, que por consequência eleva o valor da CTC.

O valor do coeficiente de correlação de Pearson ( $r$ ) é afetado pela amplitude da variação das variáveis. O valor de  $r$  tende a diminuir quando a variabilidade das variáveis é menor. Quanto maior a variabilidade das variáveis, maior a chance de detectar a correlação entre elas. Este fato pode explicar em parte os baixos valores do coeficiente de correlação, significativos ao nível de 0,1% de probabilidade, obtidos entre os atributos de solos e as covariáveis ambientais.

O uso do SRC (*stepwise* removendo a colinearidade) não resultou em grande perda para o coeficiente  $R^2$  ajustado, e permitiu consideráveis simplificações nos modelos de RLM. O método de regressão é mais eficiente quando as covariáveis não apresentam valores elevados de correlação entre si (colinearidade).

Carvalho Junior *et al.* (2014) realizaram estudo similar, em uma região de relevo movimentado, usando covariáveis ambientais derivadas do MDE (15 m de resolução) e imagens Landsat 5, e selecionando modelos preditivos de RLM para as atributos pH, MOS e Silte. Os resultados dos valores de  $R^2$  dos modelos testados por eles são relativamente mais baixos que os valores encontrados no presente estudo. Este fato indica que o uso do MDE de maior escala e resolução espacial, com o uso de informações adicionais, derivadas de dados gamaespectrométricos, resulta em modelos preditivos melhores.

Em uma avaliação final, os resultados da predição espacial usando RLM e *Random Forest* apresentam aspectos relevantes da relação solo-paisagem. Esses aspectos podem ser observados nos valores mais elevados dos teores de MOS, silte e argila nas posições de fundo de vale (região de deposição de material fino e acúmulo de umidade), nos teores mais elevados de areia nas posições de solos mais jovens (regiões de Neossolos Litólicos e Cambissolos), valores baixo de pH encontrados nas regiões de Gleissolos e Organossolos e valores CTC acompanham os resultados do teor de argila e MOS.

Os modelos *Random Forest* apresentam melhores resultados na predição de cinco atributos de solos avaliados. Com exceção do pH, o *Random Forest* apresentou maiores valores de  $R^2$  e menores RMSE na validação, se comparado com os resultados da RML. Ließ *et al.* (2012) observaram que os modelos *Random Forest* apresentam melhores resultados na predição de atributos físicas de solos (teores de areia, silte e

argila), quando comparados com o método de árvore de regressão, em um estudo realizado nos Andes Equatorianos, em uma área de aproximadamente 30 km<sup>2</sup>.

## 2.5 CONCLUSÕES

O uso de atributos de terreno, derivados do MDE de maior escala e resolução espacial (MDE 5 m), melhora os resultados da predição espacial dos atributos de solos em regiões de relevo movimentado.

Para o mapeamento digital de atributos químicos e físicos de solos, em escalas semidetalhadas, as covariáveis derivadas do MDE, são mais relevantes que as covariáveis derivadas de dados gamaespectrométricos e de imagem do satélite Landsat 8. Entretanto, a combinação de covariáveis ambientais de diferentes fontes (MDE, imagens de satélite e dados gamaespectrométricos) melhora os resultados dos modelos preditivos, com o uso de técnicas de mineração de dados.

A remoção das covariáveis colineares dos modelos de RLM possibilita a simplificação dos modelos, mantendo potencial preditivo similar aos modelos com covariáveis colineares.

Os modelos *Random Forest* são mais eficientes do que os modelos de Regressão Linear Múltipla. Os resultados mostram que a *Random Forest* é uma técnica promissora para a predição dos atributos de solos, usando covariáveis ambientais, principalmente em áreas extensas com dados escassos de solos.

## REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO JR, W.; CHAGAS, C. D. S.; LAGACHERIE, P.; FILHO, B. C.; BHERING, S. B. Evaluation of statistical and geostatistical models of digital soil properties mapping in tropical mountain regions. **Revista Brasileira de Ciência do Solo**, v. 38, p. 706-717, 2014.

CATEN, A. T.; DALMOLIN, R. S. D.; PEDRON, F. D. A.; SANTOS, M. D. L. M. Resolução espacial de um modelo digital de elevação definida pela função wavelet. **Pesq. agropec. bras.**, v. 47, n. 3, p. 449-457, 2012.

CPRM. **Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos**. Serviço Geológico de Brasil. 2010

CUNHA, A. M. **Seleção de variáveis ambientais e de algoritmos de classificação para o mapeamento digital de solos**. 2013. 121 p. (Doutorado em Solos e Nutrição de Plantas) - Departamento de Solos, Universidade Federal de Viçosa, Viçosa, MG, 2013.

EMBRAPA, Ed. **Manual de métodos de análise de solo**. Rio de Janeiro, RJ: Embrapa, p. 230, 2. ed. rev. ed. 2011.

ESRI. **ArcGIS Desktop for 10.2**. Environmental Systems Research Institute, Disponível em: < <http://www.esri.com> > Acesso em: junho de 2013.

FOX, J.; MONETTE, G. Generalized Collinearity Diagnostics. **Journal of the American Statistical Association**, v. 87, n. 417, p. 178-183, 1992.

GLOBALSOILMAP.NET. **Specifications Version 1 GlobalSoilMap.net products**. Disponível em: < <http://globalsoilmap.net/> > Acesso em: julho de 2013.

GRIMM, R.; BEHRENS, T.; MÄRKER, M.; ELSENBEER, H. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. **Geoderma**, v. 146, n. 1-2, p. 102-113, 2008.

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, n. 3-4, p. 383-398, 2005.

HEUNG, B.; BULMER, C. E.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, v. 214-215, p. 141-154, 2014.

LIEß, M.; GLASER, B.; HUWE, B. Uncertainty in the spatial prediction of soil texture. **Geoderma**, v. 170, p. 70-79, 2012.

LINDEMAN, R. H.; MERENDA, P. F.; GOLD, R. Z. **Introduction to Bivariate and Multivariate Analysis**. Glenview, IL: Scott Foresman & Co, 1980. 444 p.

MCKENZIE, N. J.; GALLANT, J. C. Digital soil mapping with improved environmental predictors and models of pedogenesis. In: LAGACHERIE, P.; MCBRATNEY, A. B., *et al.* (Eds.). **Digital Soil Mapping: An Introductory Perspective** ed. Amsterdam: Elsevier, 2007. p. 327-349. (Developments in Soil Science, 31).

MCKENZIE, N. J.; RYAN, P. J. Spatial prediction of soil properties using environmental correlation. **Geoderma**, v. 89, n. 1-2, p. 67-94, 1999.

MENDONCA-SANTOS, M. D. L.; MCBRATNEY, A. B.; MINASNY, B. Soil prediction with spatially decomposed environmental factors. In: LAGACHERIE, P.; MCBRATNEY, A. B., *et al.* (Eds.). **Digital Soil Mapping: An Introductory Perspective** ed. Amsterdam: Elsevier, 2007. p. 269-278. (Developments in Soil Science, 31).

MINASNY, B.; MCBRATNEY, A. B. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. **Geoderma**, v. 140, p. 324-336, 2007.

PRACILIO, G.; ADAMS, M. L.; SMETTEM, K. R. J.; HARPER, R. J. Determination of spatial distribution patterns of clay and plant available potassium contents in surface soils at the farm scale using high resolution gamma ray spectrometry. **Plant and Soil**, v. 282, n. 1-2, p. 67-82, 2006.

R DEVELOPMENT CORE TEAM. **The R Project for Statistical Computing**. Disponível em: < <http://www.r-project.org/> > Acesso em: novembro de 2013.

ROSSEL, R. A. V.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, v. 158, n. 1-2, p. 46-54, 2010.

ROSSEL, R. A. V.; CHEN, C. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. **Remote Sensing of Environment**, v. 115, n. 6, p. 1443-1455, 2011.

RUIZ, H. A. Incremento da exatidão da análise granulométrica do solo por meio da coleta da suspensão. **Revista Brasileira de Ciência do Solo**, v. 29, p. 297-300, 2005.

SAGA DEVELOPMENT TEAM. **System for Automated Geoscientific Analyses (SAGA)**. Disponível em: < <http://www.saga-gis.org/en/index.html> > Acesso em: junho 2014.

WILFORD, J. R.; BIERWIRTH, P. N.; CRAIG, M. A. Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. **Journal of Australian Geology & Geophysics**, v. 17, p. 201 - 216, 1997.

## **CAPÍTULO 3: COMPARAÇÃO DE MODELOS GEOESTATÍSTICOS PARA A PREDIÇÃO DOS ATRIBUTOS DE SOLOS**

### **RESUMO**

Mapas dos atributos de solos são geralmente produzidos através da interpolação de medições pontuais. A exatidão da interpolação pode ser melhorada através da incorporação de informações secundárias (covariáveis ambientais), principalmente quando os dados dos atributos de solos são escassos ou de baixa correlação espacial. Vários estudos indicam relações importantes entre os atributos de solos e covariáveis contínuas, espacialmente densas, derivadas de modelos digitais de elevação, de imagens de satélites e de espectrometria de raios gama. Assim, o objetivo deste estudo foi avaliar o desempenho dos modelos geoestatísticos de krigagem ordinária, cokrigagem e regressão-krigagem na predição espacial dos atributos de solos (pH, CTC, MOS, areia, silte e argila) incorporando informações auxiliares de covariáveis ambientais nos modelos preditivos. A krigagem ordinária apresentou os menores valores de  $R^2$  na validação cruzada, e a regressão-krigagem os maiores valores de  $R^2$  para os atributos de solos avaliados. Concluímos que os modelos geoestatísticos que incorporam informações auxiliares nas predições dos atributos de solos apresentam melhores resultados que os modelos que faz uso apenas de informações de solos, e que os modelos de regressão-krigagem apresentam grande potencial para o mapeamento digital dos atributos de solos em áreas extensas, com dados de solos escassos.

Termos para indexação: pedometria, covariáveis ambientais, krigagem, cokrigagem, regressão-krigagem.

### **COMPARISON OF GEOSTATISTICAL MODELS FOR PREDICTION OF SOIL PROPERTIES**

#### **ABSTRACT**

Soil attribute maps are usually produced by interpolation of point measurements. The accuracy of the interpolation can be improved by incorporating secondary information (environmental covariates), especially when the soil properties data is scarce or low spatial correlation. Several studies indicate important relationships between the soil properties and continuous covariates, spatially dense, derived from digital elevation models, satellite imagery and gamma-ray spectrometry. So, the objective of this study

was to evaluate the performance of geostatistical models of ordinary kriging, cokriging and regression-kriging in spatial prediction of soil properties (pH, CEC, MOS, sand, silt and clay) incorporating auxiliary information of environmental covariates in the models predictive. The ordinary kriging had the lowest  $R^2$  values in the cross-validation and regression-kriging the largest  $R^2$  values for the soil properties evaluated. We conclude that the geostatistical models that incorporate auxiliary information in the soil properties predictions show better results than the models that makes use only of soil information, and that the regression-kriging models have great potential for the digital mapping of soil properties in large areas with scarce soil data.

Index terms: pedometrics, environmental covariates, ordinary kriging, cokriging, regression-kriging.

### 3.1 INTRODUÇÃO

Mapas de variabilidade espacial dos atributos de solos de alta qualidade são necessários para a agricultura e gestão ambiental. Estes mapas são geralmente produzidos através da interpolação de medições pontuais. A exatidão da interpolação pode ser melhorada, através da escolha do método mais adequado e da incorporação de informações secundárias, covariáveis ambientais, útil na interpolação, principalmente quando os dados dos atributos de solos são escassos ou de baixa correlação espacial (Shi *et al.*, 2011).

Mapas dos atributos de solos, tais como: pH, CTC (capacidade de troca catiônica), MOS (matéria orgânica do solo), de teores de argila, silte e areia, fazem parte do conjunto mínimo de informações sobre solo, recomendado pelo consórcio Global Soil Map. Esse consórcio é uma iniciativa do grupo de trabalho de mapeamento digital de solo da União Internacional de Ciências do Solo - IUSS (Sanchez *et al.*, 2009).

Com relação às covariáveis ambientais, estudos indicam relações importantes entre os atributos de solos e covariáveis contínuas, espacialmente densas, derivadas de modelos digitais de elevação (MDE) (Florinsky *et al.*, 2002; Carvalho Junior *et al.*, 2014), de imagens de satélites (Rossel e Chen, 2011; Lu *et al.*, 2013) e de espectrometria de raios gama (Pracilio *et al.*, 2006; Wilford e Minty, 2007).

O objetivo deste estudo foi avaliar os modelos geoestatísticos de predição espacial dos atributos de solos, incorporando informações de covariáveis ambientais. Foram comparados os modelos geoestatísticos de krigagem ordinária (KO), cokrigagem

(CK) e regressão-krigagem (RK) e os efeitos do uso das covariáveis ambientais na predição dos atributos de solos.

## 3.2 MATERIAL E MÉTODOS

### 3.2.1 Área de estudo

A bacia do Rio Pongal está situada no município de Anchieta, litoral sul do estado do Espírito Santo, na região sudeste do Brasil. A área tem, aproximadamente, 140 km<sup>2</sup> e localiza-se entre as latitudes 20°41'S e 20°48'S e longitudes 40°38'W e 40°49'W (Figura 1). O clima é Aw, conforme classificação de Köppen, com estação seca no inverno e verão quente e chuvoso; a precipitação média anual varia de 1200 a 1400 mm, e a temperatura apresenta variação máxima entre 27 e 35 °C e mínima entre 17 e 22 °C.

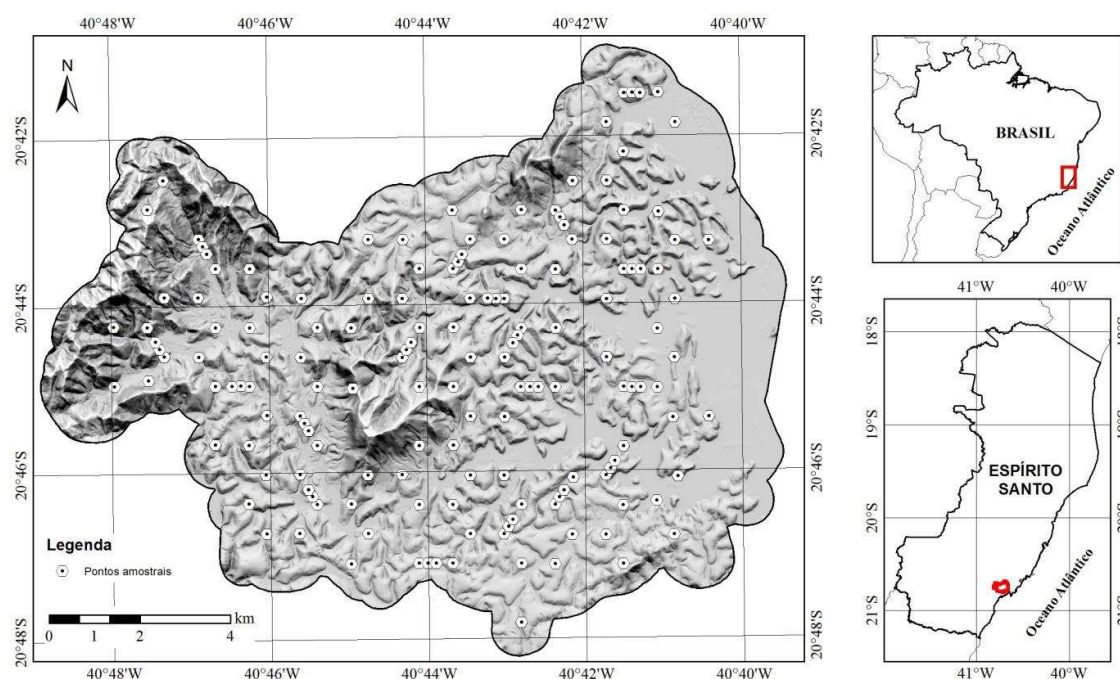


Figura 1 - Localização da área de estudo no estado do Espírito Santo, Brasil.

A pedologia da região é diversificada, com domínio de Cambissolos Háplicos Tb Distróficos nas regiões mais elevadas e íngremes do relevo, predomínio de Latossolos Amarelos Distróficos em posições intermediárias da paisagem, e Gleissolos Háplicos Tb Distróficos encaixados nas posições de fundos de vale aplainados. As classes de Latossolos Vermelho-Amarelos Distróficos, Argissolos Amarelos Distrocosos, Gleissolos Háplicos Sódicos e Neossolo Flúvico Tb Distrófico ocorrem em menor proporção na área da bacia (RADAMBRASIL, 1983).

### 3.2.2 Amostragem e análises de solos

Um total de 154 amostras foram coletadas na camada de 15 a 30 cm de profundidade do solo (GlobalSoilMap.net, 2011). Desse total, 118 amostras foram distribuídas na forma de grade regular hexagonal com 750 m de lado, e 36 amostras, coletadas em posições intermediárias na grade regular, para avaliação da variabilidade espacial dos dados em intervalos de distâncias menores que 750 m (Yfantis *et al.*, 1987; Goovaerts, 1997).

Todos os pontos de coleta de amostras foram georreferenciadas com receptor GPS no modo pós-processado com correção diferencial (DGPS), e suas respectivas coordenadas foram calculadas no sistema de projeção UTM fuso 24 sul, datum SIRGAS 2000. A distribuição dos pontos de amostragem na área de estudo é apresentada na Figura 1.

As amostras foram secas ao ar e passada na peneira de 2 mm para obtenção da TFSA (terra fina seca ao ar). Os atributos de solos: pH H<sub>2</sub>O (-log [H<sup>+</sup>]), CTC (capacidade de troca catiônica a pH 7 em cmol<sub>c</sub> dm<sup>-3</sup>), MOS (matéria orgânica do solo em dag kg<sup>-1</sup>), e os teores (dag kg<sup>-1</sup>) de areia, silte e argila, foram determinadas por análises laboratoriais de rotina, segundo Embrapa (2011) e Ruiz (2005) para a separação da fração silte.

Os valores discrepantes (*outliers*) de cada atributo de solos foram removidos usando o critério o Z-Scores maiores que três, que exclui os valores fora dos limites da média amostral mais ou menos três desvios padrões.

### 3.2.3 Covariáveis ambientais

Um conjunto com 31 covariáveis ambientais foi gerado para o uso na predição dos atributos de solos (Tabela 1). Quatorze covariáveis foram gerados a partir do Modelo Digital de Elevação (MDE), com 5 m de resolução, do levantamento aerofotogramétrico do estado do Espírito Santo, escala 1:10.000, realizado no ano de 2012. Outras seis covariáveis foram obtidas a partir dos dados gamaestrométricos do contorno radiométrico (CPRM, 2010). Imagens Landsat 8 foram usadas para obtenção de onze covariáveis adicionais.

As covariáveis derivadas do MDE são: elevação (Elev), face de exposição (Aspect), inclinação (Slope), curvatura (Curvat), perfil de curvatura (ProfCurv), multi resolução de topos aplainados (MRRTF), multi resolução de fundos de vales aplainados (MRVBF), posição de declividade média (MSP), índice de rugosidade do terreno (TRI), índice topográfico de umidade (TWI), radiação solar total (SolarRad); radiação solar

direta (DiretRad), radiação solar difusa (DiffuRad) e duração da radiação solar direta (DiretDur). Os atributos topográficos foram determinados no programa SAGA versão 2.0.8 (SAGA DEVELOPMENT TEAM, 2014) e no programa Arcgis versão 10.2 (ESRI, 2013).

Tabela 1 - Covariáveis ambientais avaliadas no mapeamento digital dos atributos de solos da bacia do Rio Pongal, fator ambiental e valores médio, mínimo e máximo.

Covariável	Fator ambiental	Média (Mín. ; Máx.)
1. Elev	relevo, clima	75,16 (0,00 ; 593,46)
2. Aspect	relevo	2,27 (-1,00 ; 6,28)
3. Slope	relevo	0,24 (0,00 ; 0,88)
4. Curvat	relevo	0,00 (-0,22 ; 0,20)
5. ProfCurv	relevo	0,00 (-0,08 ; 0,05)
6. MRRTF	relevo	0,35 (0,00 ; 4,99)
7. MRVBF	relevo	0,90 (0,00 ; 6,99)
8. MSP	relevo	0,63 (0,00 ; 0,99)
9. TRI	relevo	0,89 (0,00 ; 4,27)
10. TWI	relevo	4,05 (1,04 ; 11,13)
11. SolarRad, MWh m <sup>-2</sup>	clima	1548 (979 ; 1753)
12. DiretRad, MWh m <sup>-2</sup>	clima	1199 (664 ; 1380)
13. DiffuRad, MWh m <sup>-2</sup>	clima	349 (280 ; 386)
14. DiretDur, hora ano <sup>-1</sup>	clima	3888 (2799 ; 4330)
15. eTh, ppm	mat. origem, tempo	30,56 (-9,90 ; 168,70)
16. eU, ppm	mat. origem, tempo	3,23 (0,50 ; 10,01)
17. Kperc, %	mat. origem, tempo	0,74 (0,00 ; 3,53)
18. ThKrazao	mat. origem, tempo	48,90 (-3,56 ; 148,72)
19. UKrazao	mat. origem, tempo	4,98 (-0,22 ; 10,43)
20. UThrazao	mat. origem, tempo	0,12 (-0,01 ; 0,38)
21. B2	organismo	7926 (7315 ; 12697)
22. B3	organismo	7590 (6334 ; 12435)
23. B4	organismo	7106 (5761 ; 13033)
24. B5	organismo	13647 (5728 ; 20834)
25. B6	organismo	10885 (5290 ; 21162)
26. B7	organismo	7831 (5122 ; 14301)
27. CMI	atrib. solo	1,38 (1,02 ; 1,70)
28. IOI	atrib. solo	0,89 (0,78 ; 1,40)
29. NDVI	organismo	0,31 (-0,10 ; 0,52)
30. rB4B3	organismo, atrib.solo	0,93 (0,83 ; 1,26)
31. rB4B7	organismo, atrib.solo	0,92 (0,61 ; 1,37)

Elev: elevação; Aspect: face de exposição; Slope: inclinação; Curvat: Curvatura; ProfCurv: perfil de curvatura; MRRTF: multi resolução de topos aplainados; MRVBF: multi resolução de fundos de vales aplainados; MSP: posição de declividade média; TRI: índice de rugosidade do terreno; TWI: índice topográfico de umidade; SolarRad: radiação solar total; DiretRad: radiação solar direta; DiffuRad: radiação solar difusa; DiretDur: duração da radiação solar direta; eTh: concentração do elemento tório; eU: concentração do elemento urânio; Kperc: concentração do elemento potássio; ThKrazao: razão eTh/K; UKrazao: razão eU/K; UThrazao: razões eU/eTh; B2: bandas 2 Landsat 8; B3: bandas 3 Landsat 8; B4: bandas 4 Landsat 8; B5: bandas 5 Landsat 8; B6: bandas 6 Landsat 8; B7: bandas 7 Landsat 8; CMI: índice minerais de argila; IOI: índice óxido de ferro; NDVI: índice de vegetação por diferença normalizada; rB4B3: razão B4/B3; rB4B7: razão B4/B7.

Dos dados gamaespectrométricos foram utilizadas as seguintes covariáveis: concentração do elemento tório (eTh), concentração do elemento urânio (eU), concentração do elemento potássio (Kperc), razão eTh/K (ThKrazao), razão eU/K (UKrazao) e razões eU/eTh (UThrazao). As seis covariáveis são imagem do Levantamento Aerogeofísico do Espírito Santo, com resolução espacial de 125m (CPRM, 2010).

Dados do sensor OLI do Landsat-8, de junho de 2014, foram utilizados no terceiro conjunto de covariáveis: banda 2 - azul (B2), banda 3 - verde (B3), banda 4 - vermelho (B4), banda 5 - infravermelho próximo (B5), banda 6 - infravermelho de ondas curtas 1 (B6), banda 7 - infravermelho de ondas curtas 2 (B7), índice minerais de argila (*clay mineral index*, CMI), índice de óxido de ferro (*iron oxide index*, IOI), índice de vegetação por diferença normalizada (*normalized difference vegetation index*, NDVI), razão entre B4 e B3 (rB4B3) e razão entre B4 e B7 (rB4B7) (Cunha, 2013; Carvalho Junior *et al.*, 2014).

### **3.2.4 Modelos preditivos**

São três as abordagens para a predição espacial dos atributos de solos: a primeira utiliza o método geoestatístico de krigagem ordinária (KO), que usa apenas o valor da variável na predição. A segunda abordagem utiliza o método de cokrigagem (CK), que usa a correlação cruzada da variável com a covariável. A terceira, e última, utiliza o método híbrido de regressão-krigagem, que reúne técnicas de regressão linear múltipla (RLM) com a krigagem dos valores dos resíduos (McBratney *et al.*, 2000).

Os processamentos estatísticos e as predições foram realizados no software estatístico R (R DEVELOPMENT CORE TEAM, 2013), com os pacotes CAR (*Companion to Applied Regression*), GSTAT (*Spatial and spatio-temporal geostatistical modelling, prediction and simulation*), OUTLIERS (*Tests for outliers*), relaimpo (*Relative importance of regressors in linear models*) e RGDAL (*Bindings for the Geospatial Data Abstraction Library*).

Mais detalhes sobre os modelos preditivos utilizados podem ser obtidos nos trabalhos de Odeh *et al.* (1994), Goovaerts (1997), Vieira (2000), Hengl *et al.* (2007) e Ließ *et al.* (2012).

#### **3.2.4.1 Krigagem ordinária**

A krigagem ordinária (KO) estima localmente a média, e é aplicada em situações em que a média estacionária não é conhecida em toda a área. Caracterizada pela

condição de não viés, em que a soma dos pesos deve ser a um. O seu estimador é definido na Equação 1:

$$\hat{z}_{KO} = \sum_{i=1}^n \lambda_i z_i \quad (19)$$

em que  $\hat{z}_{KO}$  é o valor estimado na krigagem ordinária,  $z_i$  são os valores da variável nos pontos amostrados  $i$ ;  $\lambda_i$  são os pesos da krigagem, que dependem da estrutura de autocorrelação espacial da variável, e são determinados de modo que a variância do erro de predição seja minimizada.

#### 3.2.4.2 Cokrigagem

Cokrigagem (CK) é uma extensão da krigagem em que as variáveis são preditas pela utilização simultânea de suas autocorrelações e sua correlação cruzada com uma covariável. Para predição por cokrigagem é necessário que exista dependência espacial da variável e da covariável, e também que exista dependência espacial entre a variável e a covariável, que é avaliada com auxílio do semivariograma cruzado.

No caso de uma única covariável, o estimador de cokrigagem é definido na Equação 2:

$$\hat{z}_{CK} = \sum_{i=1}^n \lambda_i z_i + \sum_{j=1}^m \chi_j u_j \quad (2)$$

em que  $\hat{z}_{CK}$  é o valor estimado na cokrigagem,  $z_i$  são os valores da variável nos pontos amostrados  $i$ ;  $u_j$  são os valores da covariável;  $\lambda_i$  e  $\chi_j$  são os pesos da cokrigagem.

A seleção da covariável para o método de cokrigagem foi realizada em duas etapas. Primeiro foram selecionadas as covariáveis que apresentaram correlação significativa com a variável ( $p\text{-value} < 0,001$ ), estimada pelo coeficiente de correlação linear de Pearson. Na segunda etapa, foi selecionada a covariável com maior coeficiente de Pearson, que apresentou também correlação espacial com a variável, estimada com semivariograma cruzado.

#### 3.2.4.3 Regressão-krigagem

Regressão-krigagem (RK) é um método geoestatístico que combina a regressão da variável em função das covariáveis (como atributos de terreno, imagens de sensoriamento remoto) com krigagem dos resíduos da regressão. Assim, a abordagem RK é baseada em duas etapas principais:

- 1) A componente determinista da variável predita pode ser explicada por um modelo de RLM com índices auxiliares;

- 2) Os resíduos da regressão são krigados para descrever sua variação espacial. O resultado da krigagem do resíduo é somado à predição da RLM a partir das covariáveis.

A seleção das covariáveis preditoras dos modelos de RLM foi feita combinando o método Stepwise, que usa o AIC (*Akaike information criterion*) na escolha do modelo, com o fator de inflação da variância para o diagnóstico da colinearidade (Fox e Monette, 1992) e a medida da importância relativa para modelos lineares (Lindeman *et al.*, 1980), para remover as covariáveis colineares de menor importância relativa.

### 3.2.5 Validação cruzada (autovalidação)

A validação cruzada *leave-one-out*, também conhecida como *Jackknife*, foi o procedimento utilizado para avaliar o erro das estimativas (incertezas) dos modelos geoestatísticos (Vieira, 2000). O desempenho dos modelos foi avaliado pelos seguintes indicadores:

- $R^2$  (coeficiente de determinação): obtido na regressão linear entre valores preditos e observados (Equação 3).

$$R^2 = 1 - \frac{\sum_{i=1}^n [\hat{z}(x_i) - z(x_i)]^2}{\sum_{i=1}^n [z(x_i) - \bar{z}]^2} \quad (3)$$

- RMSE (*Root Mean Square Error* - raiz quadrada do erro quadrático médio): menores valores de RMSE indicam que os valores estimados pela interpolação estão mais próximos dos valores medidos (Equação 4).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\hat{z}(x_i) - z(x_i)]^2} \quad (4)$$

- RMSS (*Root Mean Square Standardized* - raiz quadrada do erro quadrático padronizado): valores próximos de um são desejáveis (Equação 5).

$$RMSS = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[ \frac{(\hat{z}(x_i) - z(x_i))}{\sigma(x_i)} \right]^2} \quad (5)$$

Nas equações 3, 4 e 5,  $\hat{z}$  é o valor predito,  $z$  é o valor observado nos pontos amostrais  $x_i$ ,  $\bar{z}$  é a média dos valores observados, e  $\sigma(x_i)$  é o desvio padrão do erro da predição.

### 3.3 RESULTADOS

#### 3.3.1 Estatística descritiva dos dados dos atributos de solos

Os resultados da análise descritiva de todos os atributos estudados (Figura 1) mostraram que a variável silte apresenta a maior variabilidade, seguida pela variável CTC, MOS e areia, com base no coeficiente de variação (CV). Na outra ponta, o pH apresentou o menor coeficiente de variação entre as variáveis analisadas.

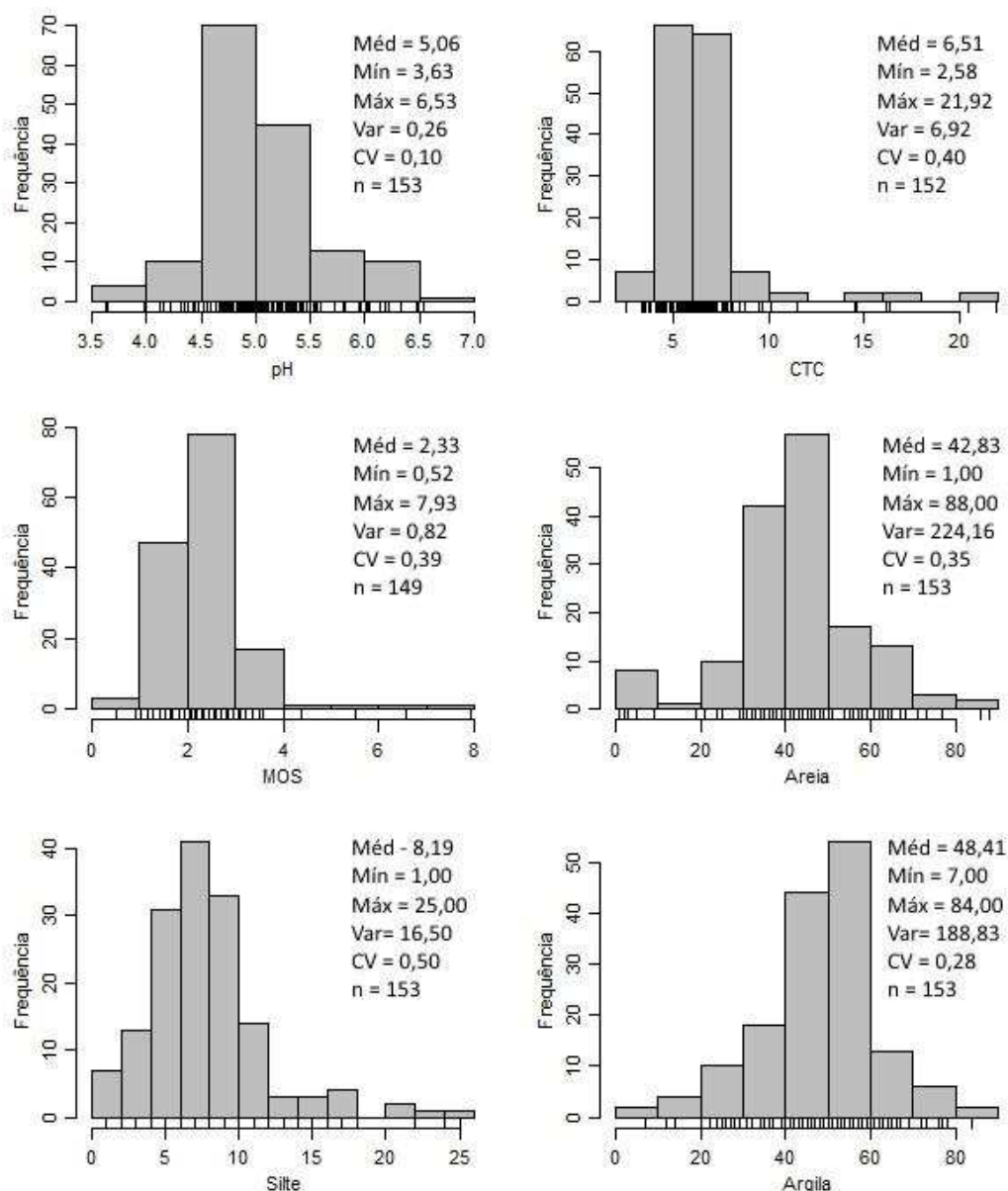


Figura 2 - Histograma e estatística descritiva dos atributos de solos, na camada de 15 a 30 cm, da bacia do Rio Pongal. pH: pH em água ( $-\log[H^+]$ ); CTC: capacidade de troca catiônica a pH 7 ( $\text{cmol}_c \text{dm}^{-3}$ ); MOS: matéria orgânica do solo ( $\text{dag kg}^{-1}$ ); Areia ( $\text{dag kg}^{-1}$ ); Silte ( $\text{dag kg}^{-1}$ ); Argila ( $\text{dag kg}^{-1}$ ). Méd: média; Mín: mínimo; Máx: máximo; Var: variância; CV: coeficiente de variação; n: número de observações.

### 3.3.2 Covariáveis preditoras

A variável pH não apresentou correlação linear significativa ( $p\text{-value} < 0,001$ ) com as covariáveis ambientais avaliadas. As covariáveis que apresentam coeficiente de correlação linear significativo com a variável areia, não apresentaram correlação espacial com a mesma. Desta forma, não foi possível prever o pH e a areia por cokrigagem (Tabela 2).

Tabela 2 - Covariáveis selecionadas para predição, por cokrigagem, dos atributos de solos da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal.

Variável <sup>(1)</sup>	Covariável correlacionada <sup>(2)</sup>	Dependência espacial <sup>(3)</sup>	Covariável selecionada
pH	Não	Não	Não
CTC	Aspect, Slope, MRRTF, MRVBF, MSP, TRI, TWI, DiretDur	Sim	TWI
MOS	Slope, MRVBF, MSP, TRI, TWI, DiffuRad, DiretDur	Sim	TWI
Areia	ProfCurv, MRVBF, TWI, DiretDur	Não	Não
Silte	Aspect, MRVBF, TWI	Sim	MRVBF
Argila	Curvat, ProfCurv, DiretDur	Sim	DiretDur

<sup>(1)</sup> pH: pH em água; CTC: capacidade de troca catiônica a pH 7,0; MOS: matéria orgânica do solo; Areia: teor de areia; Silte: teor de silte; Argila: teor de argila. <sup>(2)</sup> Covariável correlacionada ao nível de 0,1% ( $p\text{-value} < 0,001$ ); Aspect: face de exposição; Slope: inclinação; Curvat: Curvatura; ProfCurv: perfil de curvatura; MRRTF: multi resolução de topos aplainados; MRVBF: multi resolução de fundos de vales aplainados; MSP: posição de declividade média; TRI: índice de rugosidade do terreno; TWI: índice topográfico de umidade; DiffuRad: radiação solar difusa; DiretDur: duração da radiação solar direta. <sup>(3)</sup> Dependência espacial entre a variável e a covariável.

O modelo de regressão linear múltipla (RLM) da variável CTC apresentou o maior coeficiente de determinação da regressão, com valor superior a 0,50. Para as demais variáveis, o modelo de RLM apresentaram valores de  $R^2$  entre 0,32 e 0,36 (Tabela 3).

Tabela 3 - Covariáveis selecionadas para predição, por regressão-krigagem, de atributos de solos da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal.

Variável <sup>(1)</sup>	Covariáveis selecionadas <sup>(2)</sup>	R <sup>2</sup> (RLM)
pH	UKrazao + Curvat + MSP + TRI + TWI	0,26
CTC	Slope + ProfCurv + TWI	0,53
MOS	ThKrazao + IsB2 + TWI	0,33
Areia	IsNDVI + ProfCurv + MRVBF + DiretDur	0,34
Silte	UKrazao + IsIOI + Elev + MRVBF + DiretRad + DiffuRad	0,39
Argila	ThKrazao + Slope + ProfCurv + MRRTF + TWI + DiretDur	0,33

<sup>(1)</sup> pH: pH em água; CTC: capacidade de troca catiônica a pH 7,0; MOS: matéria orgânica do solo; Areia: teor de areia; Silte: teor de silte; Argila: teor de argila. <sup>(2)</sup> UKrazao: razão eU/K; ThKrazao: razão eTh/K; IsB2: bandas 2 Landsat 8; IsNDVI: *normalized difference vegetation index*; IsIOI: *iron oxide index*; Elev: elevação; Slope: inclinação; Curvat: Curvatura; ProfCurv: perfil de curvatura; MRRTF: Multi-Resolution Ridge Top Flatness; MRVBF: Multi-Resolution Valley Bottom Flatness; MSP: Mean Position of Declivity; TRI: Terrain Ruggedness Index; TWI: Topographic Wetness Index; DiretRad: radiação solar direta; DiffuRad: radiação solar difusa; DiretDur: duração da radiação solar direta. R<sup>2</sup>(RLM): Coeficiente de determinação da regressão linear múltipla.

### 3.3.3 Predição espacial dos atributos de solos

Para a krigagem ordinária da variável pH, o ajuste do semivariograma empírico foi realizado com o modelo esférico. Para todas as demais variáveis (CTC, MOS, areia, silte e argila) o ajuste foi realizado com o modelo exponenciais, independente do modelo geoestatístico utilizado (Tabela 4).

Os valores de alcance foram relativamente próximos, entres os modelos ajustados para as variáveis CTC, MOS e Argila, variando entre 350 e 650 m aproximadamente. A variável pH apresentou grande variação entre os valores de alcance da função de krigagem ordinária, que ultrapassou 1500 m e de regressão-krigagem, que foi inferior a 220 m.

Tabela 4 - Parâmetros dos modelos geostatísticos e da validação cruzada das predições espaciais dos atributos de solos, da camada 15 a 30 cm de profundidade, da bacia do Rio Pongal.

Variável <sup>(1)</sup>	Função <sup>(2)</sup>	Modelo <sup>(3)</sup>	C0	C1	a	R <sup>2</sup>	RMSE	RMSS
pH (-log[H <sup>+</sup> ])	KO	Esf	0,21	0,05	1518	0,00	0,38	1,03
	CK	NA	-	-	-	-	-	-
	RK	Exp	0,00	0,18	199	0,19	0,35	1,01
CTC (cmol <sub>c</sub> dm <sup>-3</sup> )	KO	Exp	0,00	6,93	547	0,00	1,58	1,19
	CK	Exp	0,19	6,81	616	0,41	1,37	1,19
	RK	Exp	0,00	3,22	510	0,48	1,23	1,04
MOS (dag kg <sup>-1</sup> )	KO	Exp	0,09	0,70	565	0,04	0,57	1,13
	CK	Exp	0,12	0,55	640	0,32	0,51	1,12
	RK	Exp	0,10	0,40	431	0,23	0,52	1,05
Areia (dag kg <sup>-1</sup> )	KO	Exp	0,00	209,10	298	0,01	10,53	1,10
	CK	NA	-	-	-	-	-	-
	RK	Exp	0,00	149,70	286	0,28	9,36	1,02
Silte (dag kg <sup>-1</sup> )	KO	Exp	0,00	14,74	361	0,03	2,83	1,11
	CK	Exp	4,57	6,64	755	0,23	2,55	1,09
	RK	Exp	0,00	9,95	280	0,30	2,42	1,03
Argila (dag kg <sup>-1</sup> )	KO	Exp	0,00	181,64	433	0,01	10,47	1,17
	CK	Exp	2,28	179,30	433	0,08	9,90	1,15
	RK	Exp	0,00	124,71	353	0,26	8,83	1,01

<sup>(1)</sup> pH: pH em água; CTC: capacidade de troca catiônica a pH 7,0; MOS: matéria orgânica do solo; Areia: teor de areia; Silte: teor de silte; Argila: teor de argila. <sup>(2)</sup> KO: krigagem ordinária; CK: cokrigagem; RK: regressão-krigagem. <sup>(3)</sup> Esf: esférico; Exp: exponencial; NA: modelo não ajustado. C0: efeito pepita; C1: contribuição; a: alcance (m); R<sup>2</sup>: coeficiente de determinação; RMSE: *Root Mean Square Error*; RMSS: *Root Mean Square Standardized*.

O resultado da krigagem ordinária para a variável pH apresentou pequena variação nos valores preditos, que ficaram no intervalo de 4,75 a 5,35, e a média variância da predição foi de 0,25. A regressão-krigagem apresentou maior variação no resultados da predição, com valores no intervalo de 3,54 a 6,66. Neste caso, a média da variância da predição foi de 0,19 (Figura 3).

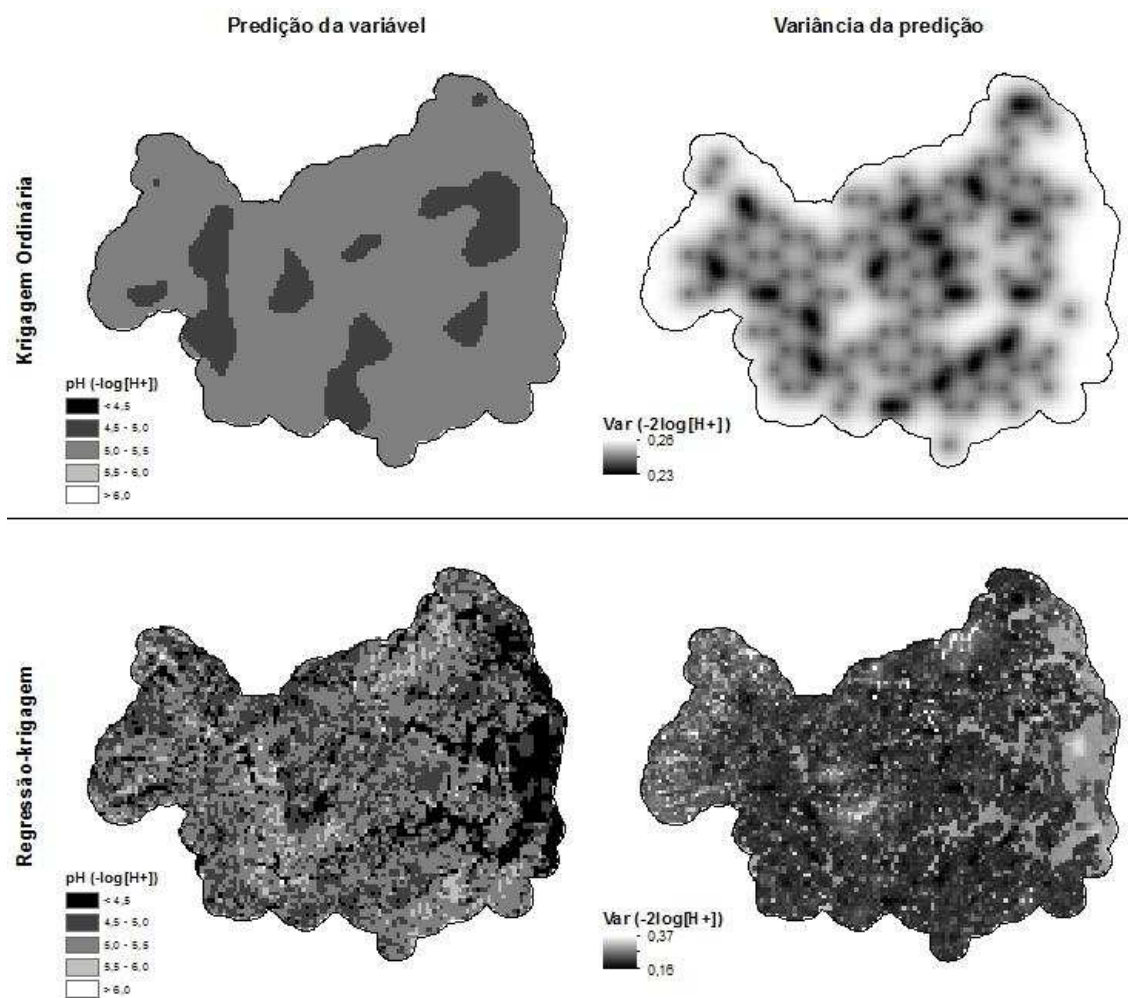


Figura 3 - Mapa da predição espacial da variável pH (pH H<sub>2</sub>O em -log[H<sup>+</sup>]) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária e regressão-krigagem.

Para a variável CTC, o resultado da krigagem ordinária apresentou valores preditos variando entre 2,8 e 19,7 cmol<sub>c</sub> dm<sup>-3</sup>, e a média da variância da predição foi de 4,99. Os resultados da cokrigagem são semelhantes aos da krigagem ordinária, com valores preditos variando entre 2,1 e 19,4 cmol<sub>c</sub> dm<sup>-3</sup>, e média da variância da predição de 4,98. Os valores preditos na regressão-krigagem ficaram no intervalo de 0,6 a 22,4 cmol<sub>c</sub> dm<sup>-3</sup>, com valor de 2,59 para a média da variância da predição (Figura 4).

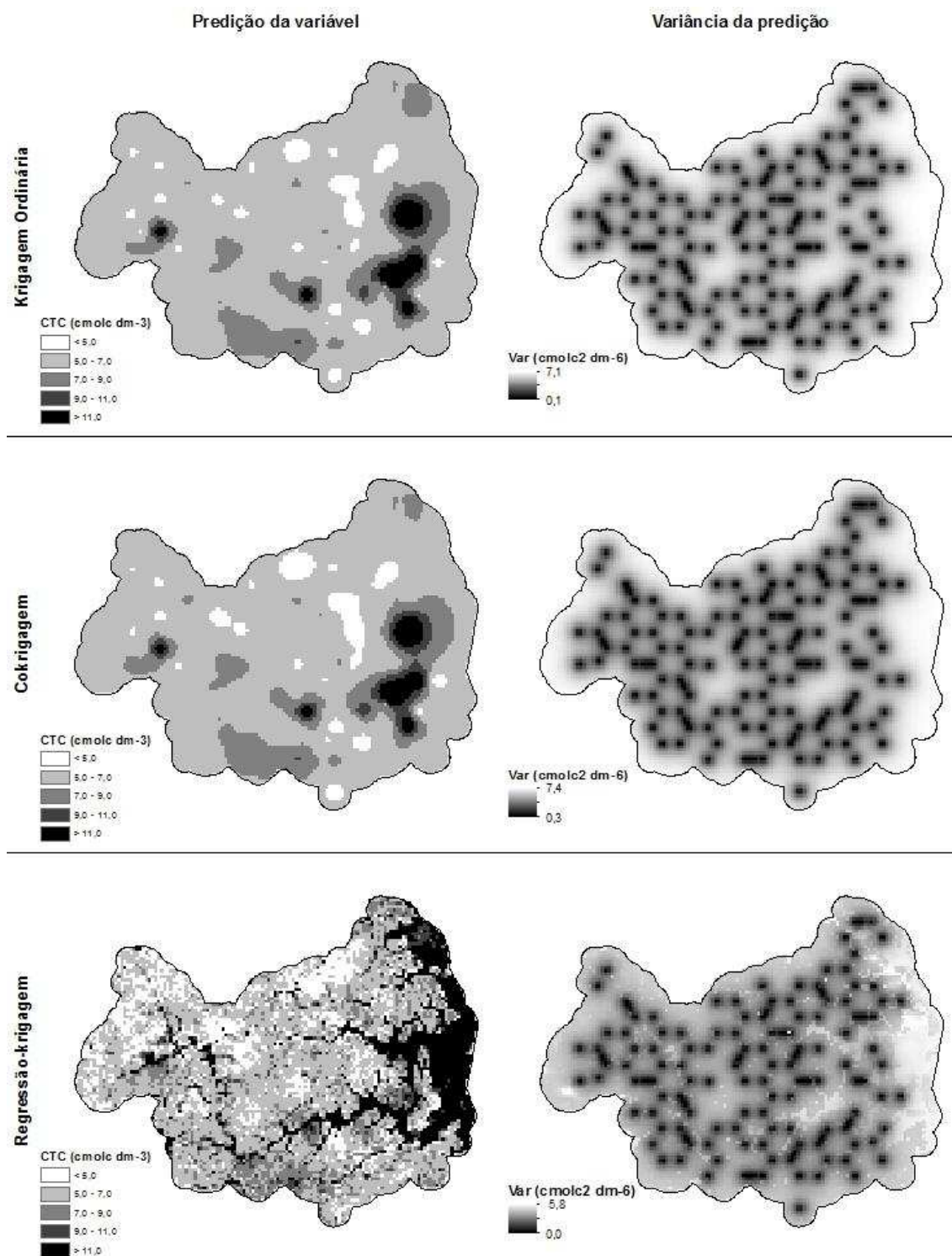


Figura 4 - Mapa da predição espacial da variável CTC (capacidade de troca catiônica a pH 7,0 em cmol<sub>c</sub> dm<sup>-3</sup>) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária, cokrigagem e regressão-krigagem.

O resultado da krigagem ordinária da variável MOS apresentou valores preditos variando entre 0,88 a 6,69 dag kg<sup>-1</sup>, e a média da variância da predição foi de 0,61. A cokrigagem apresentou valores preditos variando no intervalo de 1,15 a 6,63 dag kg<sup>-1</sup>, e a média da variância da predição foi de 0,60. A regressão-krigagem apresentou valores

preditos variando no intervalo de 0,60 a 6,77 dag kg<sup>-1</sup>. Neste caso, a média da variância da predição foi de 0,46, valor que é inferior a variância dos dados (Figura 5).

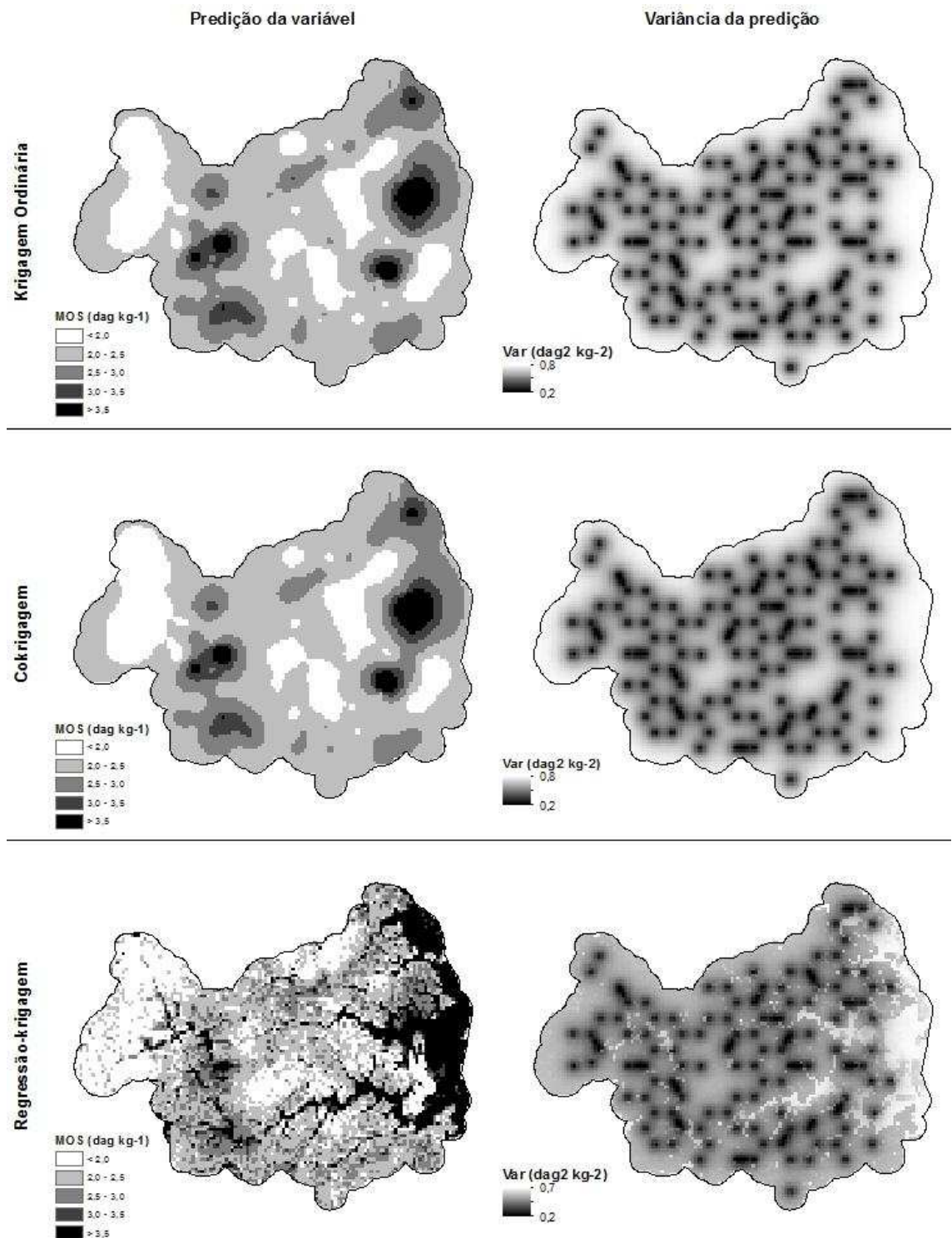


Figura 5 - Mapa da predição espacial da variável MOS (Matéria Orgânica do Solo em dag kg<sup>-1</sup>) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária, cokrigagem e regressão-krigagem.

Para a variável Areia, o resultado da krigagem ordinária apresentou valores preditos variando entre 7 e 79 dag kg<sup>-1</sup>, e a média da variância da predição foi de 185. Os valores preditos na regressão-krigagem ficaram no intervalo de -8 a 112 dag kg<sup>-1</sup>, com valor de 139 para a média da variância da predição (Figura 6).

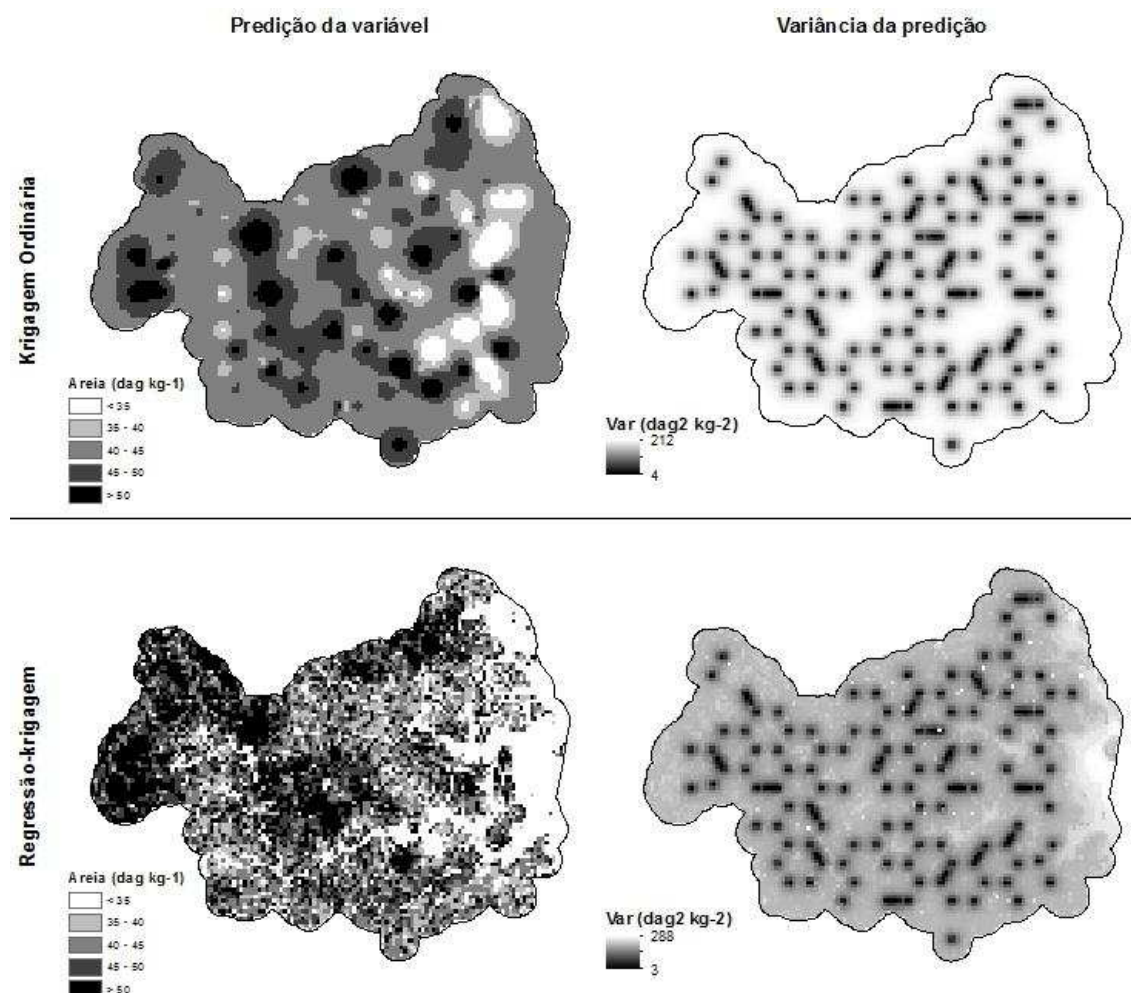


Figura 6 - Mapa da predição espacial da variável Areia (teor de areia em dag kg<sup>-1</sup>) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária e regressão-krigagem.

Os valores preditos para a variável Silte, usando krigagem ordinária, variaram entre 1,9 a 22,9 dag kg<sup>-1</sup>, e a média da variância da predição foi de 12,4. A cokrigagem apresentou valores preditos variando no intervalo de 3,3 a 21,4 dag kg<sup>-1</sup>, e a média da variância da predição foi de 12,0. A regressão-krigagem apresentou valores preditos variando no intervalo de -2,8 a 23,8 dag kg<sup>-1</sup>, e a média da variância da predição foi de 9,9 (Figura 7).

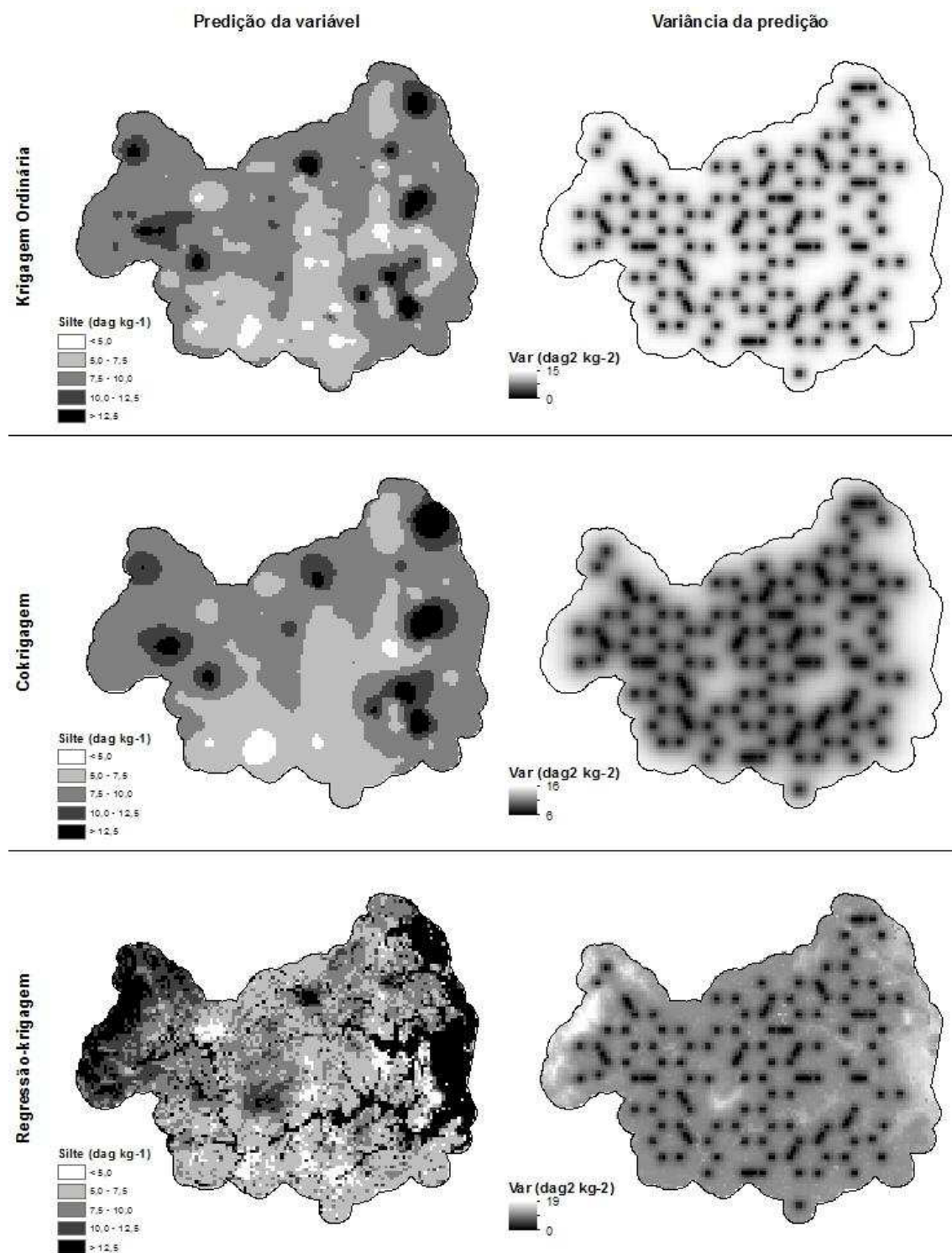


Figura 7 - Mapa da predição espacial da variável Silte (teor de silte em  $\text{dag kg}^{-1}$ ) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária, cokrigagem e regressão-krigagem.

O resultado da krigagem ordinária da variável Argila apresentou valores preditos variando entre 11 e 80  $\text{dag kg}^{-1}$ , e a média da variância da predição foi de 144. A cokrigagem apresentou valores preditos variando no intervalo de 13 a 79  $\text{dag kg}^{-1}$ , e a média da variância da predição foi de 146. A regressão-krigagem apresentou valores

preditos variando no intervalo de -11 a 91 dag kg<sup>-1</sup>, e a média da variância da predição foi de 113 (Figura 8).

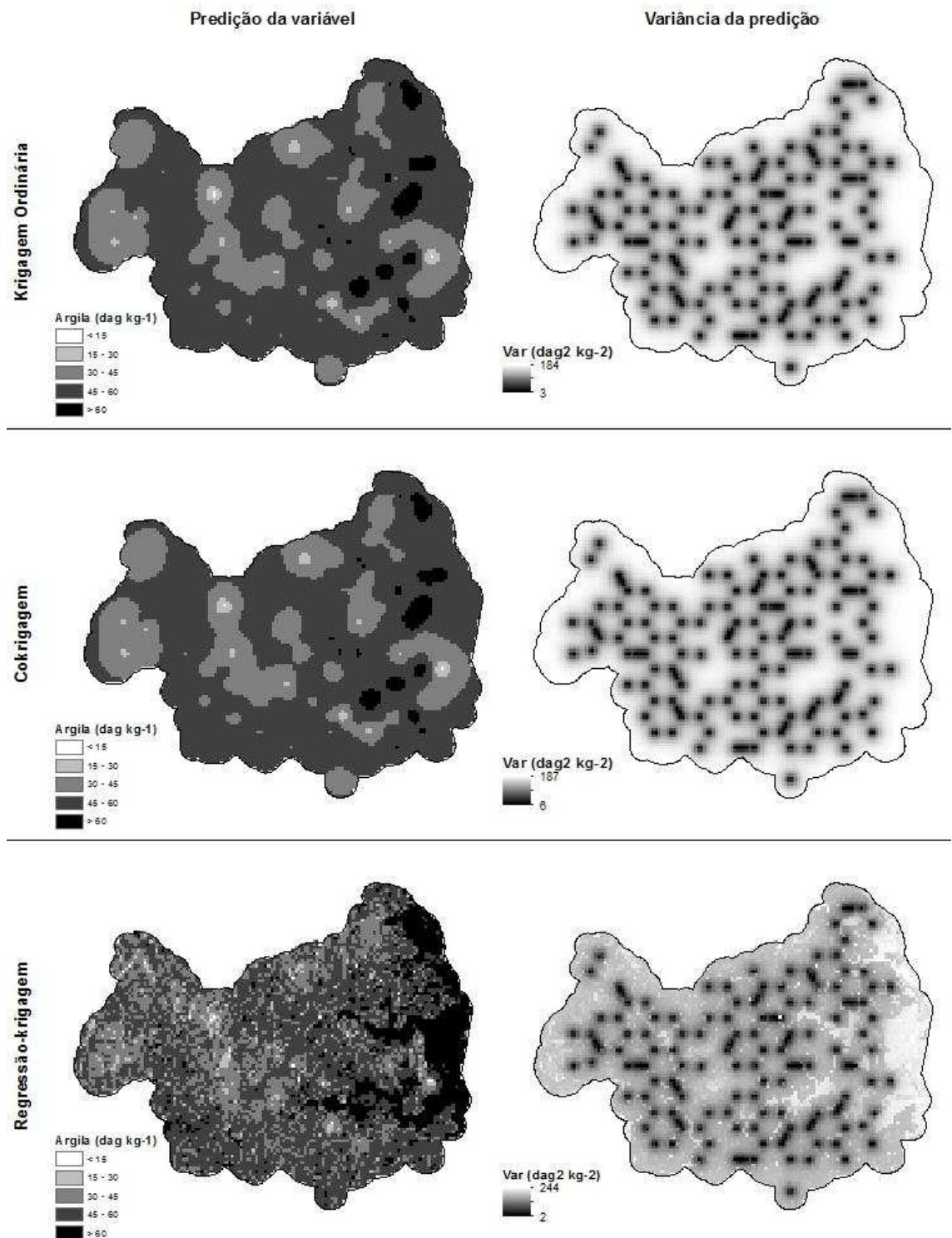


Figura 8 - Mapa da predição espacial da variável Argila (teor de argila em dag kg<sup>-1</sup>) e variância da predição, da camada 15 a 30 cm de profundidade, da Bacia do Rio Pongal, por krigagem ordinária, cokrigagem e regressão-krigagem.

### 3.4 DISCUSSÃO

O desempenho da krigagem ordinária (KO) foi muito próximo ao desempenho da cokrigagem (CK), considerando os mapas da predição. No entanto, os indicadores de exatidão da validação cruzada ( $R^2$ , RMSE e RMSS) mostram que o desempenho da CK, que usa a informação da covariável ambiental para predição dos valores da variável em locais não amostrados, é melhor que o desempenho da KO, que usa apenas os valores dos atributos de solos.

A cokrigagem é mais exigente que a krigagem, pois é necessária a existência de dependência espacial de cada uma das variáveis (autocorrelação espacial) e a existência de correlação cruzada entre elas (Goovaerts, 1999). Satisfazer as condições requeridas pela cokrigagem representa a maior dificuldade para aplicação do método, restringindo sua aplicação, como nos casos das variáveis pH e Areia. O fraco desempenho da CK, com resultados relativamente próximos aos da KO, está relacionado com a baixa correlação espacial entre a variável e a covariável utilizada na predição. O uso de covariáveis de mais alto grau de correlação espacial com os atributos de solo produz melhores resultados na predição espacial por cokrigagem.

Os resultados dos indicadores de exatidão da validação cruzada ( $R^2$ , RMSE e RMSS) da regressão-krigagem (RK) foram melhores que os demais métodos geoestatísticos de krigagem ordinária e cokrigagem. Apenas para a variável MOS, os indicadores  $R^2$  e RMSE da RK foram piores que os indicadores da CK, contudo o RMSS da RK foi menor que o da CK, comprovando o melhor desempenho da RK.

Qi-yong *et al.* (2014) observaram que a RK apresentou melhores resultados nos indicadores de validação, se comparados com a krigagem ordinária e a regressão linear múltipla, para predição espacial da MOS, em uma região de topografia complexa, com área de 10,73 km<sup>2</sup>, no sudoeste da China.

Em uma avaliação qualitativa dos mapas, os resultados da predição espacial por regressão-krigagem apresentam aspectos relevantes da relação solo-paisagem. Esses aspectos podem ser observados pelos: 1) valores mais elevados dos teores de MOS, silte e argila nas posições de fundo de vale, região de deposição de material fino e acúmulo de umidade; 2) teores mais elevados de areia nas posições de solos pouco desenvolvidos, regiões de Neossolos Litólicos e Cambissolos; 3) valores baixo de pH encontrados nas regiões de Gleissolos e Organossolos; e 4) valores CTC acompanhando os valores teor de argila e MOS. Esta é a maior vantagem dos métodos híbridos, que permitem utilizar relações solo-paisagem na predição dos atributos de solos.

Os resultados do coeficiente de determinação da validação cruzada ( $R^2$ ) para regressão-krigagem foram maiores que os observado no estudo realizado por Carvalho Junior *et al.* (2014), que encontram  $R^2$  variando no intervalo de 0,09 a 0,18 nas predições das variáveis argila, silte, areia e pH e MOS, para a camada do solo de 15 a 30 cm, em um estudo similar.

Uma desvantagem da regressão-krigagem é que ela produz valores discrepantes em alguns casos, como da predição de valores negativos para as variáveis areia, silte e argila, e valores superiores a 100 para a variável areia. Estes valores discrepantes são resultado de ruídos gerados na regressão linear múltipla, e podem ser corrigidos, após a predição, identificando a localização dos valores discrepantes e recalculando-os a partir dos valores dos pixels vizinhos.

### 3.5 CONCLUSÕES

Para a predição dos atributos de solos em escala semidetalhadas de mapeamento, os métodos geoestatísticos, que incorporam informações auxiliares (covariáveis ambientais) nos modelos preditivos, apresentam melhores resultados que os métodos que fazem uso apenas das informações das variáveis.

O modelo geoestatístico de regressão-krigagem é mais eficiente do que os modelos de krigagem ordinária e cokrigagem. A regressão-krigagem, que permite usar as relações solo-paisagem na predição, apresenta grande potencial para o mapeamento digital dos atributos de solos.

### REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO JR, W.; LAGACHERIE, P.; CHAGAS, C. D. S.; CALDERANO FILHO, B.; BHERING, S. B. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma*, v. 232-234, p. 479-486, 2014.

CPRM. **Programa Levantamentos Geológicos Básicos do Brasil**. Píuma. Folha SF.24-V-A-VI. Estado do Espírito Santo. Escala 1:100.000. Brasília, DF: DNPM/CPRM: 140 p. 1993.

CPRM. **Relatório final do levantamento e processamento dos dados magnetométricos e gamaespectrométricos**. Serviço Geológico de Brasil. 2010

CUNHA, A. M. **Seleção de variáveis ambientais e de algoritmos de classificação para o mapeamento digital de solos**. 2013. 121 p. (Doutorado em Solos e Nutrição de Plantas) - Departamento de Solos, Universidade Federal de Viçosa, Viçosa, MG, 2013.

- EMBRAPA, Ed. **Manual de métodos de análise de solo**. Rio de Janeiro, RJ: Embrapa, p. 230, 2. ed. rev. ed. 2011.
- ESRI. **ArcGIS Desktop for 10.2**. Environmental Systems Research Institute, Disponível em: < <http://www.esri.com> > Acesso em: junho de 2013.
- FLORINSKYA, I. V.; EILERS, R. G.; MANNING, G. R.; FULLER, L. G. Prediction of soil properties by digital terrain modelling. **Environmental Modelling & Software**, v. 17, p. 295-311, 2002.
- FOX, J.; MONETTE, G. Generalized Collinearity Diagnostics. **Journal of the American Statistical Association**, v. 87, n. 417, p. 178-183, 1992.
- GLOBALSOILMAP.NET. **Specifications Version 1 GlobalSoilMap.net products**. Disponível em: < <http://globalsoilmap.net/> > Acesso em: julho de 2013.
- GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York, NY: Oxford University Press, 1997. 483 p.
- GOOVAERTS, P. Geostatistics in soil science: state-of-the-art and perspectives. **Geoderma**, v. 89, p. 1-45, 1999.
- HENGL, T.; HEUVELINK, G. B. M.; ROSSITER, D. G. About regression-kriging: From equations to case studies. **Computers & Geosciences**, v. 33, n. 10, p. 1301-1315, 2007.
- LIEß, M.; GLASER, B.; HUWE, B. Uncertainty in the spatial prediction of soil texture. **Geoderma**, v. 170, p. 70-79, 2012.
- LINDEMAN, R. H.; MERENDA, P. F.; GOLD, R. Z. **Introduction to Bivariate and Multivariate Analysis**. Glenview, IL: Scott Foresman & Co, 1980. 444 p.
- LU, P.; WANG, L.; NIU, Z.; LI, L.; ZHANG, W. Prediction of soil properties using laboratory VIS-NIR spectroscopy and Hyperion imagery. **Journal of Geochemical Exploration**, v. 132, p. 26-33, 2013.
- MCBRATNEY, A. B.; ODEH, I. O. A.; BISHOP, T. F. A.; DUNBAR, M. S.; SHATAR, T. M. An overview of pedometric techniques for use in soil survey. **Geoderma**, v. 97, p. 293-327, 2000.
- ODEH, I. O. A.; MCBRATNEY, A. B.; CHITTLEBOROUGH, D. J. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. **Geoderma**, v. 63, p. 197-214, 1994.
- PRACILIO, G.; ADAMS, M. L.; SMETTEM, K. R. J.; HARPER, R. J. Determination of spatial distribution patterns of clay and plant available potassium contents in surface soils at the farm scale using high resolution gamma ray spectrometry. **Plant and Soil**, v. 282, n. 1-2, p. 67-82, 2006.
- QI-YONG, Y.; ZHONG-CHENG, J.; WEN-JUN, L.; HUI, L. Prediction of soil organic matter in peak-cluster depression region using kriging and terrain indices. **Soil & Tillage Research**, v. 144, p. 126-132, 2014.

R DEVELOPMENT CORE TEAM. **The R Project for Statistical Computing**. Disponível em: < <http://www.r-project.org/> > Acesso em: novembro de 2013.

RADAMBRASIL. **Levantamento de Recursos Naturais Volume 32**. Folhas SF 23/24 Rio de Janeiro/Vitória: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Rio de Janeiro, RJ: Projeto RADAMBRASIL: 780 p. 1983.

ROSSEL, R. A. V.; CHEN, C. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. **Remote Sensing of Environment**, v. 115, n. 6, p. 1443-1455, 2011.

RUIZ, H. A. Incremento da exatidão da análise granulométrica do solo por meio da coleta da suspensão. **Revista Brasileira de Ciência do Solo**, v. 29, p. 297-300, 2005.

SAGA DEVELOPMENT TEAM. **System for Automated Geoscientific Analyses (SAGA)**. Disponível em: < <http://www.saga-gis.org/en/index.html> > Acesso em: junho de 2014.

SANCHEZ, P. A.; AHAMED, S.; CARRE, F.; HARTEMINK, A. E.; HEMPEL, J.; HUISING, J.; LAGACHERIE, P.; MCBRATNEY, A. B.; MCKENZIE, N. J.; MENDONÇA-SANTOS, M. L.; MINASNY, B.; MONTANARELLA, L.; OKOTH, P.; PALM, C. A.; SACHS, J. D.; SHEPHERD, K. D.; VAGEN, T. G.; VANLAUWE, B.; WALSH, M. G.; WINOWIECKI, L. A.; ZHANG, G. L. Digital soil map of the world. **Science**, v. 325, n. 5941, p. 680-681, 2009.

SHI, W.; LIU, J.; DU, Z.; STEIN, A.; YUE, T. Surface modelling of soil properties based on land use information. **Geoderma**, v. 162, p. 347-357, 2011.

VIEIRA, S. R. Geoestatística em estudos da variabilidade espacial do solo In: NOVAIS, R. F.; ALVARES V., V. H., *et al.* (Eds.). **Tópicos em Ciência do Solo**. 1 ed. Viçosa, MG: SBCS, v. 1, 2000. p. 1-54.

WILFORD, J.; MINTY, B. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. In: LAGACHERIE, P.; MCBRATNEY, A. B., *et al.* (Eds.). **Digital Soil Mapping: An Introductory Perspective** ed. Amsterdam: Elsevier, 2007. p. 207-218. (Developments in Soil Science, 31).

## CONCLUSÕES GERAIS

O uso da geoestatística para a predição espacial de atributos de solos representa uma alternativa viável para o mapeamento digital de solos em escala semidetalhadas. A densidade de amostras exigidas e, conseqüentemente, os custos com a amostragem e as análises laboratoriais, são similares aos custos do mapeamento convencional (clássico).

Em situações de escassez de dados de solo, o uso de covariáveis ambientais melhora a exatidão das predições espaciais dos atributos de solos. As técnicas de mineração de dados possibilitam a seleção das covariáveis mais relevantes para os modelos preditivos.

Entre as covariáveis ambientais comumente utilizadas na predição, os atributos de terreno apresentam correlações significativas com os atributos de solos em regiões de relevo complexo. Nestas regiões, os MDEs de maior escala e resolução espacial apresentam os melhores resultados da predição espacial dos atributo de solos.

Os modelos *Random Forest* foram mais eficientes que os modelos de Regressão Linear Múltipla, comparando os resultados obtidos na validação. Vantagens como a facilidade de aplicação e a robustez do método de predição, fazem da *Random Forest* uma técnica promissora para a predição dos atributos de solos.

A regressão-krigagem, técnica híbrida que combinam RLM e geoestatística, apresenta o melhor desempenho na predição dos atributos de solo, quando comparada com a krigagem ordinária e a cokrigagem. A regressão-krigagem permite usar as relações solo-paisagem na predição, que melhora a exatidão dos modelos.

E apesar do fraco desempenho dos modelos preditivos avaliados, que apresentaram coeficiente de determinação ( $R^2$ ) inferiores a 0,60, a predição espacial dos atributos de solo vem evoluindo rapidamente. O emprego de novas técnicas e o uso de covariáveis ambientais (auxiliares) de melhor qualidade e maior grau de correlação com os atributos de solos, pode melhorar o desempenho dos modelos preditivos.