

MÉTODO DO HIPERCUBO LATINO CONDICIONADO PARA A AMOSTRAGEM DE SOLOS NA PRESENÇA DE COVARIÁVEIS AMBIENTAIS VISANDO O MAPEAMENTO DIGITAL DE SOLOS⁽¹⁾

Waldir de Carvalho Júnior⁽²⁾, César da Silva Chagas⁽²⁾, Alexandre Muselli⁽³⁾, Helena Saraiva Koenow Pinheiro⁽⁴⁾, Nilson Rendeiro Pereira⁽⁵⁾ & Silvio Barge Bhering⁽²⁾

RESUMO

A amostragem é uma das etapas mais importantes dos levantamentos de solos. No entanto, os esquemas de amostragem utilizados nos levantamentos convencionais têm se evidenciado inadequados para o mapeamento digital de solos, pois podem comprometer os resultados e, além disso, não possibilitam a realização de análises estatísticas. Este estudo teve por objetivo avaliar o método de amostragem do hipercubo latino condicionado (cLHS, sigla em inglês), na presença de covariáveis ambientais (elevação, declividade, curvatura e mapa de uso e cobertura do solo), em comparação com a amostragem aleatória, na alocação de 100 pontos amostrais, buscando maior representatividade das características ambientais da bacia do rio Guapi-Macacu. O desempenho dos métodos foi avaliado pela análise qualitativa dos histogramas de frequência e das análises estatísticas pelos testes F, T de Student e Kolmogorov-Smirnov (K-S), para cada covariável. Os resultados apresentaram que os pontos selecionados pelo método cLHS possuíam distribuição geográfica mais adequada do que aqueles obtidos pela amostragem aleatória. Além disso, o método cLHS preservou mais a distribuição de frequência das covariáveis contínuas do que a amostragem aleatória; para covariável categórica uso e cobertura do solo os métodos foram equivalentes. Os testes estatísticos confirmaram o melhor desempenho do método cLHS, cujas amostras não diferiram estatisticamente da bacia. Entretanto, a amostragem aleatória apresentou diferença estatística para com a bacia, para todas as covariáveis

⁽¹⁾ Trabalho submetido em 17 de maio de 2013 e aprovado em 31 de outubro de 2013.

⁽²⁾ Pesquisador A, Embrapa Solos. Rua Jardim Botânico, 1024. CEP 22460-000 Rio de Janeiro (RJ). E-mail: waldir.carvalho@embrapa.br, cesar.chagas@embrapa.br, chagas.rj@gmail.com, silvio.bhering@embrapa.br

⁽³⁾ Pesquisador Assistente, Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Av. Prof. Almeida Prado, 532, Cid. Universitária. CEP 05508-901 São Paulo (SP). E-mail: alexandre.muselli@gmail.com

⁽⁴⁾ Doutoranda do Departamento de Solos, Universidade Federal Rural do Rio de Janeiro. BR 465, km 47. CEP 23890-000 Seropédica (RJ). E-mail: lenask@gmail.com

⁽⁵⁾ Pesquisador B, Embrapa Solos. E-mail: nilson.pereira@embrapa.br

contínuas para pelo menos um dos testes utilizados. Assim, o método cLHS pode ser considerado como um método satisfatório para seleção de locais de amostragem em áreas heterogêneas similares as deste estudo, visando a utilização no mapeamento digital de solos.

Termos de indexação: esquema de amostragem, levantamento de solos, modelo digital de elevação.

SUMMARY: CONDITIONED LATIN HYPERCUBE METHOD FOR SOIL SAMPLING IN THE PRESENCE OF ENVIRONMENTAL COVARIATES FOR DIGITAL SOIL MAPPING

Soil sampling is one of the most important stages in soil surveys. However, in conventional soil surveys, sampling schemes have proven to be inadequate for digital soil mapping because they may affect the results obtained and, in addition, they do not allow statistical estimations to be performed. This study attempts to evaluate the conditioned Latin Hypercube Sampling (cLHS) method with four environmental covariates (elevation, slope, curvature and land cover / land use map), in comparison to random sampling, in allocating 100 sample points, seeking a more representative description of the environmental characteristics of the Macacu watershed area. The performance of the methods was evaluated by qualitative assessment of histograms of distribution, and statistical analyses by the F test, Student T Test and Kolmogorov-Smirnov test (K-S) for each covariable. The sampling points selected by the cLHS method showed better geographic distribution than those obtained by random sampling. Furthermore, the cLHS method better maintained the frequency distribution of the continuous covariables than random sampling did, and the categorical covariable of land cover / land use showed no significance differences between the cLHS and random methods. Statistical analysis confirmed the better performance of the cLHS method compared to the random method, and the cLHS samples showed no significance differences from the watershed data. In contrast, the random samples showed significance differences from the watershed data for all the continuous covariables for at least one of the statistical tests used. Thus, the cLHS method is a satisfactory method for selecting points for soil sampling in heterogeneous areas, like those of the present study, for the purpose of use in digital soil mapping.

Index terms: sampling design, soil survey, digital elevation model.

INTRODUÇÃO

Uma importante etapa dos levantamentos de solos consiste nos trabalhos de campo, em que são feitas as observações e a coleta de amostras em pontos específicos, transectos ou áreas previamente estabelecidas; normalmente é a etapa mais onerosa desse tipo de estudo (Webster & Olivier, 1990; Embrapa, 1995). Nos levantamentos convencionais, os locais de amostragem são selecionados subjetivamente pelo pedólogo para suportar seu modelo mental preditivo de ocorrência dos solos (Brungard & Boettinger, 2010), levando em consideração conceitos como intensidade ou tipo do levantamento, variabilidade espacial, escala do mapeamento, objetivos e composição das unidades de mapeamento (amostragem livre).

Segundo Hengl (2003), os esquemas de amostragem utilizados nos levantamentos de solos convencionais são direcionados, não aleatórios e não fornecem estimativas estatísticas representativas da totalidade da área de estudo. Além disso, não existe consenso quanto à densidade de observações ideal e, como outros fatores além do nível de detalhamento influenciam o

número final de amostras, principalmente a complexidade de ocorrência dos solos em cada região, torna-se difícil estabelecer parâmetros rígidos por nível de levantamento (Embrapa, 1995).

Diferentemente do mapeamento convencional, no mapeamento digital de solos muita atenção tem sido dada para otimização da amostragem, utilizando diferentes métodos como: amostragem aleatória (Gessler et al., 1995; Howell et al., 2007); amostragem aleatória estratificada (McKenzie & Ryan, 1999; Roecker & Thompson, 2010); análise multivariada (Lesch et al., 1995; Hengl et al., 2003); lógica fuzzy (Brus et al., 2007; Zhu et al., 2008); e geoestatística (Brus & Heuvelink, 2007; Vašát et al., 2010).

De acordo com Minasny & McBratney (2006), um esquema de amostragem com base na utilização de dados auxiliares deve ser capaz de capturar toda a variação de valores das diferentes covariáveis auxiliares utilizadas. Para tanto, esses autores propuseram o método cLHS, que é de uso livre, e que se caracteriza como ferramenta robusta para a alocação de pontos amostrais pela utilização de um conjunto de covariáveis contínuas e, ou, categóricas. O cLHS utiliza como base o método de amostragem

do hipercubo latino (LHS, sigla em inglês), proposto por McKay et al. (1979).

Minasny & McBratney (2006) compararam o método cLHS com a amostragem aleatória e a aleatória estratificada (Brus et al., 2007) e demonstraram que sua utilização com as covariáveis declividade, índice topográfico combinado (CTI, sigla em inglês), NDVI (índice de vegetação por diferença normalizada, sigla em inglês) e uso do solo representou a distribuição dessas covariáveis, com um número de amostras relativamente pequeno, melhor do que os outros dois métodos utilizados. Resultados semelhantes foram alcançados por Minasny & McBratney (2007), ao compararem o método cLHS com a amostragem aleatória simples, amostragem aleatória estratificada, amostragem por componentes principais (Hengl et al., 2003) e o método cLHS com incorporação das coordenadas espaciais.

Segundo Brungard & Boettinger (2010), a determinação do número ideal de amostras, quando se utiliza o método cLHS, ainda é um problema a ser investigado; sendo assim, buscaram determinar esse número ideal utilizando seis covariáveis (declividade, inverso do índice CTI, aspecto, índice de vegetação - *Soil Adjusted Vegetation Index* -, uso do solo e geologia). Os resultados evidenciaram que o método cLHS representou adequadamente a distribuição das covariáveis utilizadas e que o número de amostras recomendadas, para áreas com condições ambientais similares, deve ser de aproximadamente 0,05 a 0,1 % do número total de *pixels* da área a ser amostrada.

Thomas et al. (2012) utilizaram o método cLHS para determinar locais para amostragem em uma bacia nas Filipinas com base na utilização de sete covariáveis (bandas 1, 2, 3 e 4 do Landsat MSS, elevação, plano e perfil de curvatura). Neste estudo, os 20 locais derivados do método cLHS forneceram uma distribuição adequada da variação geográfica e ambiental da área estudada, sendo esses melhores do que os obtidos com a amostragem aleatória e a amostragem em *grid*, com os quais o método cLHS foi comparado. As principais limitações apontadas foram a necessidade de revisão das covariáveis utilizadas, de modo a serem mais representativas das características pedogenéticas da bacia e a natureza de “caixa-preta” que inibe a interação com pedólogos experientes.

Para Roudier et al. (2012), embora o método cLHS tenha se tornado uma ferramenta interessante no mapeamento digital de solos, esse frequentemente produz um esquema amostral que resulta em planos de implementação dispendiosa. Por essa razão, propuseram melhoria no método pela incorporação de restrições operacionais. O custo para se alcançar cada ponto na paisagem a partir das estradas foi avaliado usando atributos topográficos e a cobertura do terreno. Esse custo foi então usado no processo de otimização do método para penalizar os pontos que eram difíceis ou impossíveis de serem alcançados no campo.

No Brasil, a utilização do método cLHS foi relatada apenas no trabalho de Santos et al. (2008). Sendo assim, este estudo teve como objetivo avaliar esse método de amostragem para a inferência de pontos amostrais na presença de atributos do terreno derivados de um modelo digital de elevação e dados de um sensor remoto orbital (Landsat 5), visando maior representatividade das características ambientais da bacia dos Rios Guapi-Macacu para sua utilização no mapeamento digital de classes de solos.

MATERIAL E MÉTODOS

A área de estudo compreende a bacia hidrográfica dos Rios Guapi-Macacu (1.250,78 km²), pertencente à região hidrográfica da Baía de Guanabara, RJ, e está localizada entre as coordenadas UTM 7.488.481 e 7.526.005 m de latitude sul e 699.292 e 752.193 m de longitude oeste, *Datum* Córrego Alegre, Zona 23 Sul.

A área da bacia se estende das escarpas serranas às planícies costeiras, passando por colinas, maciços e tabuleiros costeiros, abrangendo uma diversidade de ambientes onde ocorrem ecossistemas diversos como florestas, campos de altitude, restingas e mangues. O material de origem dos solos é formado predominantemente por gnaisse do Complexo Paraíba do Sul; migmatitos e granitos referentes ao Complexo Rio Negro, Suíte Serra dos Órgãos e Suíte Rio de Janeiro; corpos intrusivos do Complexo Alcalino de Rio Bonito (sienitos e nefelina-sienitos); e Sedimentos do Quaternário de origem continental e transicional/marinho, divididos em Sedimentos Paludais, Sedimentos Marinhos, Sedimentos Litorâneos e Sedimentos Fluviais (Silva & Cunha, 2001).

Os solos predominantes na área da bacia são: Argissolos Amarelos e Vermelho-Amarelos, Cambissolos Háplicos, Gleissolos Háplicos, Melânicos, Tiomórficos e Sállicos, Latossolos Amarelos e Vermelho-Amarelos, Neossolos Flúvicos e Litólicos, e Planossolos Háplicos (Embrapa, 2003).

Covariáveis ambientais

Um modelo digital de elevação (MDE) da bacia, com resolução espacial de 30 m, obtido com a utilização do módulo *Topo to Raster* do ArcGIS Desktop 10, a partir de curvas de nível, hidrografia e pontos contidos nas cartas topográficas do IBGE/DSG (folhas Itaboraí, Itaipava, Nova Friburgo, Petrópolis, Teresópolis e Rio Bonito), todos na escala de 1:50.000, além de pontos cotados extraídos do SRTM (Pinheiro et al., 2012), foi utilizado para gerar, no *Spatial Analyst* do ArcGIS Desktop 10 (ESRI, 2010), os atributos do terreno elevação, da declividade e da curvatura do terreno (Gallant & Wilson, 2000), identificados por ocasião dos trabalhos de campo (Pinheiro, 2012) e considerados como os principais fatores do relevo que controlam a distribuição dos solos na bacia. Esses atributos do

terreno têm sido reconhecidos como os mais efetivos para a realização de levantamento de solos de média escala (Chagas et al., 2010; Carvalho Júnior et al., 2011).

Dados sobre uso e cobertura do solo (Embrapa, 2008) foram também utilizados. As classes estabelecidas neste estudo foram: afloramento rochoso, água, área agrícola, área urbana, campo de altitude, mangue, pastagem e vegetação florestal em diferentes estádios de conservação.

Embora a bacia apresente diferentes litologias, não se verificou, durante os trabalhos de campo (Pinheiro, 2012), diferenças significativas entre os solos derivados de gnaisses, granitos, migmatitos e rochas alcalinas nas porções mais elevadas da bacia, sendo as diferenças mais acentuadas verificadas apenas entre os solos derivados desses materiais, daqueles originados dos sedimentos do quaternário que ocupam os setores mais baixos da área. Por essa razão, optou-se por não utilizar o mapa geológico como covariável ambiental, já que as diferenças citadas poderiam ser identificadas pelos atributos do terreno elevação e declividade.

Métodos de amostragem e análises estatísticas

O método cLHS é um procedimento de amostragem aleatória estratificada que fornece uma eficiente forma de amostragem de variáveis a partir de suas distribuições multivariadas (Minasny & McBratney, 2006). O método cLHS funciona da seguinte forma: dado K variáveis com $X_1, \dots, X_k$, sendo a faixa de variação de cada uma, X é dividido em n prováveis intervalos iguais (estratos); para cada variável uma amostra aleatória é tomada para cada estrato. As amostras obtidas para cada variável por estrato são confrontadas umas com as outras, de forma aleatória ou seguindo alguma regra previamente especificada. A regra neste caso é de que a amostragem possa refletir a mesma representação dos estratos para todas as variáveis consideradas. Ao final, tem-se um número de amostras que cobrem os n estratos para todas as variáveis analisadas. Nessa técnica, são necessárias menos amostras para se obter uma distribuição mais representativa das características ambientais da área estudada. O método garante que cada variável está representada de maneira completa por meio de seus estratos. Uma descrição mais pormenorizada sobre o método cLHS pode ser obtida em Minasny & McBratney (2006) e Roudier et al. (2012).

Para facilitar a coleta das amostras e tornar essa etapa menos onerosa em termos de custo e de tempo, visto que a acessibilidade a algumas áreas da bacia (elevada declividade do terreno, vegetação densa nas escarpas serranas e, ou, ausência de vias de acesso) poderia dificultar essa etapa, foram incorporadas algumas restrições operacionais, conforme Roudier et al. (2012), utilizando como regra de restrição espacial um *buffer* de 100 m de largura ao longo das principais estradas e vias de acesso, além da exclusão das áreas

urbanas, das unidades de conservação e dos corpos d'água, configurando assim a área efetiva de estudo. Do mesmo modo, foi usada como referência para comparação dos resultados a área da bacia, excluindo-se as áreas das Unidades de Conservação (UCs), evitando assim resultados muito discrepantes que poderiam comprometer a avaliação.

Após a aplicação das restrições, os dados das covariáveis ambientais para a área efetiva (elevação, declividade, curvatura e uso e cobertura do solo) foram convertidos para o formato de arquivo-texto e utilizados como dados de entrada no programa cLHS (Minasny & McBratney, 2006). Uma vez geradas as bases para o processo, procedeu-se a seleção de 100 pontos amostrais, utilizando peso dos dados igual a 1,0, peso das correlações igual a 0,5 e número de interações igual a 20.000, conforme especificações do programa (Minasny & McBratney, 2006). Como ainda não existem critérios definitivos para estabelecer o número ideal de pontos amostrais, quando se utiliza o método cLHS, optou-se por selecionar 100 pontos, considerados a princípio suficientes para captar toda a variabilidade das covariáveis ambientais e, consequentemente, dos solos da bacia, embora estudo de Brungard & Boettinger (2010) aponte como número ideal de amostras entre 0,05 e 0,1 % do total de células da área estudada.

De maneira a permitir a comparação com o método cLHS, foram também selecionados, empregando-se as mesmas regras de restrição espacial utilizadas nesse método, 100 pontos aleatoriamente (amostragem aleatória), utilizando-se o pacote estatístico do R (R Development Core Team, 2013).

Com o auxílio de um Sistema de Posicionamento Global (GPS), nos pontos selecionados pelo método cLHS, foram descritos e coletados perfis e amostras extras, conforme Santos et al. (2005). As amostras de solo coletadas foram submetidas às análises, seguindo os procedimentos normalmente utilizados nos levantamentos de solos convencionais, disponíveis em Embrapa (1997). Os solos foram classificados no 4º nível categórico, segundo o Sistema Brasileiro de Classificação de Solos (Embrapa, 2006).

Os testes estatísticos F , utilizado para comparar a variância entre amostras; T de Student, para comparação de médias; e Kolmogorov-Smirnov ($K-S$), usado para comparar a localização, dispersão e forma da distribuição entre amostras, foram empregados para verificar a existência ou não de diferenças significativas entre as amostras selecionadas pelo método cLHS e pela amostragem aleatória com as amostras da bacia e da área efetiva, com relação às covariáveis contínuas (Teetor, 2011). Para construir essa análise, foi utilizado o pacote estatístico R (R Development Core Team, 2013). Para todos os testes foi observado o valor p para um nível de confiança de 95 %, com valores maiores que 0,05 indicando a inexistência de diferenças entre os conjuntos testados.

RESULTADOS E DISCUSSÃO

Os histogramas de frequência das covariáveis ambientais utilizadas e a estatística descritiva das covariáveis contínuas são apresentados na figura 1. De acordo com Minasny & McBratney (2006, 2007) e Thomas et al. (2012), uma análise mais acurada da distribuição das diferentes covariáveis ambientais nos diferentes métodos de amostragem pode ser obtida por meio de seus histogramas.

De modo geral, verificou-se, na comparação com a bacia, que a área efetiva preservou satisfatoriamente a distribuição das covariáveis contínuas elevação (Figura 1a), declividade (Figura 1b) e curvatura (Figura 1c), sendo os percentuais das classes muito parecidos. O uso e a cobertura do solo (covariável categórica) evidenciaram uma distribuição de classes diferente entre a bacia e a área efetiva, simplesmente pelo fato de as regras restritivas terem excluído as áreas dos afloramentos de rocha, água e áreas urbanas (Figura 1d). Por essa razão, apenas as classes culturas, pastagem e floresta foram contempladas na área efetiva. A classe mangue, relacionada com solos muito mal-drenados que sofrem influência marinha, também não foi contemplada em razão da falta de acessos.

Os resultados na seleção dos pontos amostrais com a utilização do método cLHS e a amostragem aleatória, assim como o mapa digital de solos no 2º nível categórico do SiBCS (Embrapa, 2006), obtido com a utilização da abordagem por redes neurais artificiais (Pinheiro, 2006), a partir dos pontos coletados com o método cLHS, são apresentados na figura 2. No quadro 1, são apresentados os resultados com a aplicação dos testes F, T de Student e de Kolmogorov-Smirnov para as covariáveis contínuas.

Verifica-se na figura 2 que os pontos selecionados pelo método cLHS apresentaram melhor distribuição geográfica do que os obtidos na amostragem aleatória, que evidencia tendência em agrupar pontos em alguns setores da bacia, conforme verificado por Minasny & McBratney (2006), contrariando a percepção geral de que uma boa amostragem requer boa distribuição geográfica.

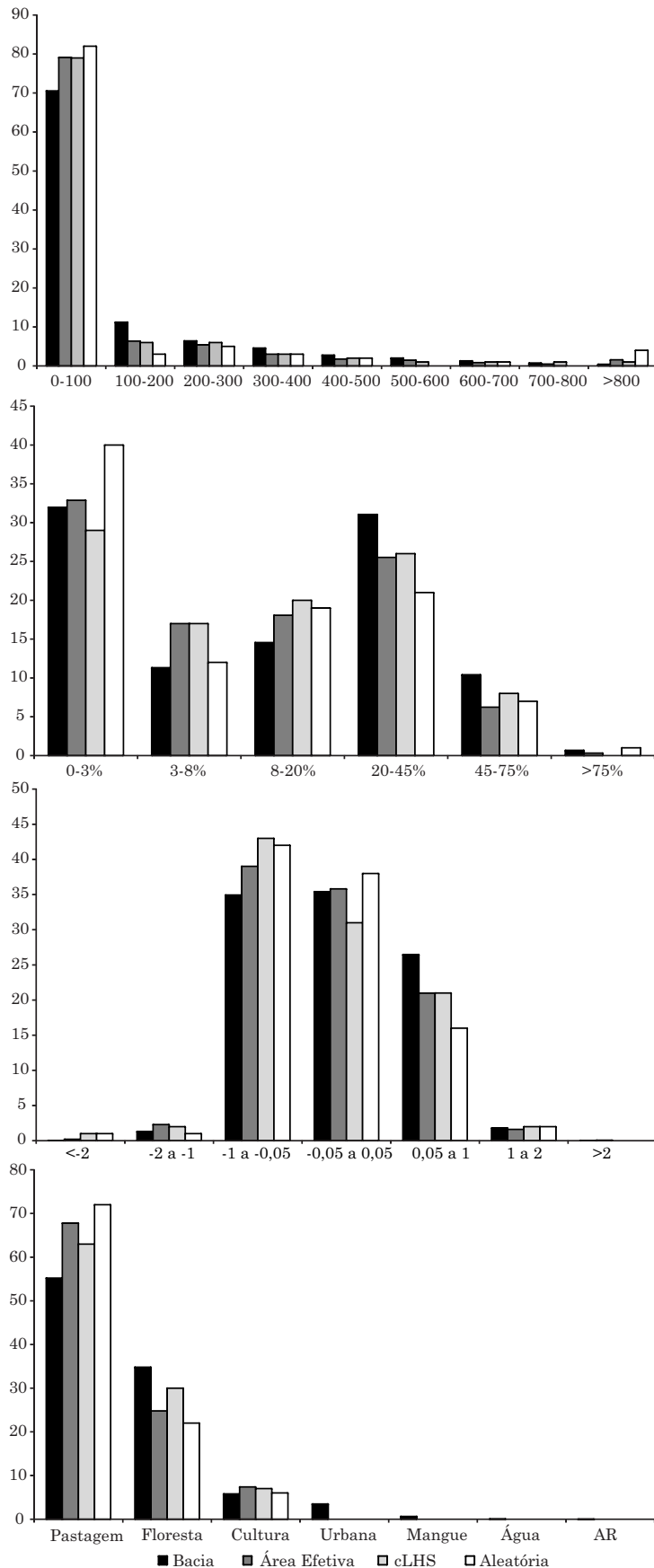
A distribuição de frequência dos valores das covariáveis ambientais apresentada na figura 1 demonstrou que o método cLHS preservou a distribuição das covariáveis contínuas melhor do que a amostragem aleatória, na comparação com a área efetiva e com a bacia, do mesmo modo como obtido por Minasny & McBratney (2006, 2007) e Thomas et al. (2012). Em relação à distribuição de frequência da covariável categórica uso e cobertura do solo, os dois métodos evidenciaram comportamento similar.

No que se refere à elevação, o histograma da distribuição dos valores obtidos pelo método cLHS apresentou que apenas 1 % das amostras foram coletadas em altitudes maiores do que 800 m, enquanto na amostragem aleatória esse percentual

foi de 4 %; entretanto, nos intervalos entre 500 e 600 m e 700 e 800 m nenhuma amostra foi coletada. Em ambos os casos, o número de amostras escolhidas (100) pode ter sido insuficiente para permitir melhor representação dessas áreas. No entanto, esses estratos de elevação representam pequena parcela em relação às áreas efetiva e da bacia, não comprometendo a representatividade da amostragem (Figura 1a). De acordo com os testes estatísticos realizados, verificou-se que para essa covariável as amostras selecionadas pelo método cLHS não possuem diferenças significativas para variância, média e distribuição com relação tanto à bacia quanto à área efetiva (Quadro 1). As amostras coletadas aleatoriamente possuem variância e distribuição estatisticamente diferentes da bacia, bem como variância estatisticamente diferente dos dados da área efetiva. Esses resultados corroboram com a análise qualitativa da distribuição de frequência pelo histograma, que apontou o método cLHS como mais adequado para representar tanto a bacia quanto a área efetiva.

Com relação à declividade, a distribuição da frequência das amostras selecionadas pelo método cLHS foi similar à distribuição verificada para a área efetiva, enquanto a amostragem aleatória tendeu a sobreamostrar as áreas com declividade inferior a 3 % e a subamostrar as áreas entre 3 e 8 % e 20 e 45 % (Figura 1b). Conforme verificado por Minasny & McBratney (2007), notou-se redução do intervalo dos valores para essa covariável, com maior ênfase no método cLHS (0 a 69 % contra 0 a 131 % na área efetiva). Essa redução se deveu ao fato de que as áreas com declividade superior a 69 % perfizeram apenas 1,50 % da área da bacia e a 0,63 % da área efetiva; como no método cLHS, o número de sorteios foi diretamente proporcional à probabilidade de ocorrência de cada estrato, esse foi desconsiderado. Essa mesma redução também foi verificada na amostragem aleatória, porém com menor magnitude (0 a 82 %).

A análise da declividade evidenciou que também não existe diferença significativa, com relação à variância, média e distribuição, entre as amostras obtidas pelo método cLHS, a bacia e a área efetiva (Quadro 1). As análises estatísticas também confirmaram a similaridade verificada entre os histogramas de frequência tanto da bacia quanto da área efetiva com as amostras alcançadas pelo método cLHS (Figura 1b). Por sua vez, as amostras obtidas aleatoriamente apresentaram diferenças significativas para a média e a distribuição, quando comparadas à bacia. Embora tenha sido verificada alguma dissimilaridade entre os histogramas de frequência da declividade das amostras aleatórias e a área efetiva (Figura 1b), nenhuma diferença estatística foi observada, o que pode estar indicando que a bacia e a área efetiva apresentaram diferenças estatísticas para essa covariável (Quadro 1). Mesmo assim, o método cLHS também se evidenciou mais adequado do que a amostragem aleatória para representar a área da bacia.



a) Elevação (m)

	n	Média	DP
Bacia	1527427	110,6	159,9
Área efetiva	299923	95,1	169,4
cLHS	100	98,4	178,9
Aleatória	100	108,7	223,5

n: n° de amostras; DP: Desvio padrão.

b) Declividade (%)

	n	Média	DP
Bacia	1527427	19,4	19,1
Área efetiva	299923	15,2	16,6
cLHS	100	16,2	16,7
Aleatória	100	15,0	18,2

n: n° de amostras; DP: Desvio padrão.

c) Curvatura

	n	Média	DP
Bacia	1527427	-0,007	0,398
Área efetiva	299923	-0,069	0,420
cLHS	100	0,089	0,449
Aleatória	100	0,077	0,434

n: n° de amostras; DP: Desvio padrão.

d) Uso e cobertura do solo

Figura 1. Histogramas das covariáveis ambientais utilizadas na bacia dos Rios Guapi-Macacu.

Tanto o método cLHS quanto a amostragem aleatória preservaram de modo satisfatório a distribuição da frequência da curvatura em relação à área efetiva, destoando ligeiramente em relação à bacia (Figura 1c). De modo geral, a amostragem pelo método cLHS superestimou ligeiramente a classe entre -1 e -0,05, subestimou as áreas de curvatura planar (-0,05 e 0,05) e manteve o percentual de distribuição para o intervalo

entre 0,05 e 1. Na amostragem aleatória, também houve ligeiro aumento para a classe de -1 e -0,05 e redução para a classe entre 0,05 e 1. As classes <-1 e >1 praticamente não tiveram seus percentuais influenciados em ambos os métodos de amostragem.

Os testes estatísticos evidenciaram que não existem diferenças significativas entre as amostras obtidas pelo método cLHS, a área efetiva e a bacia

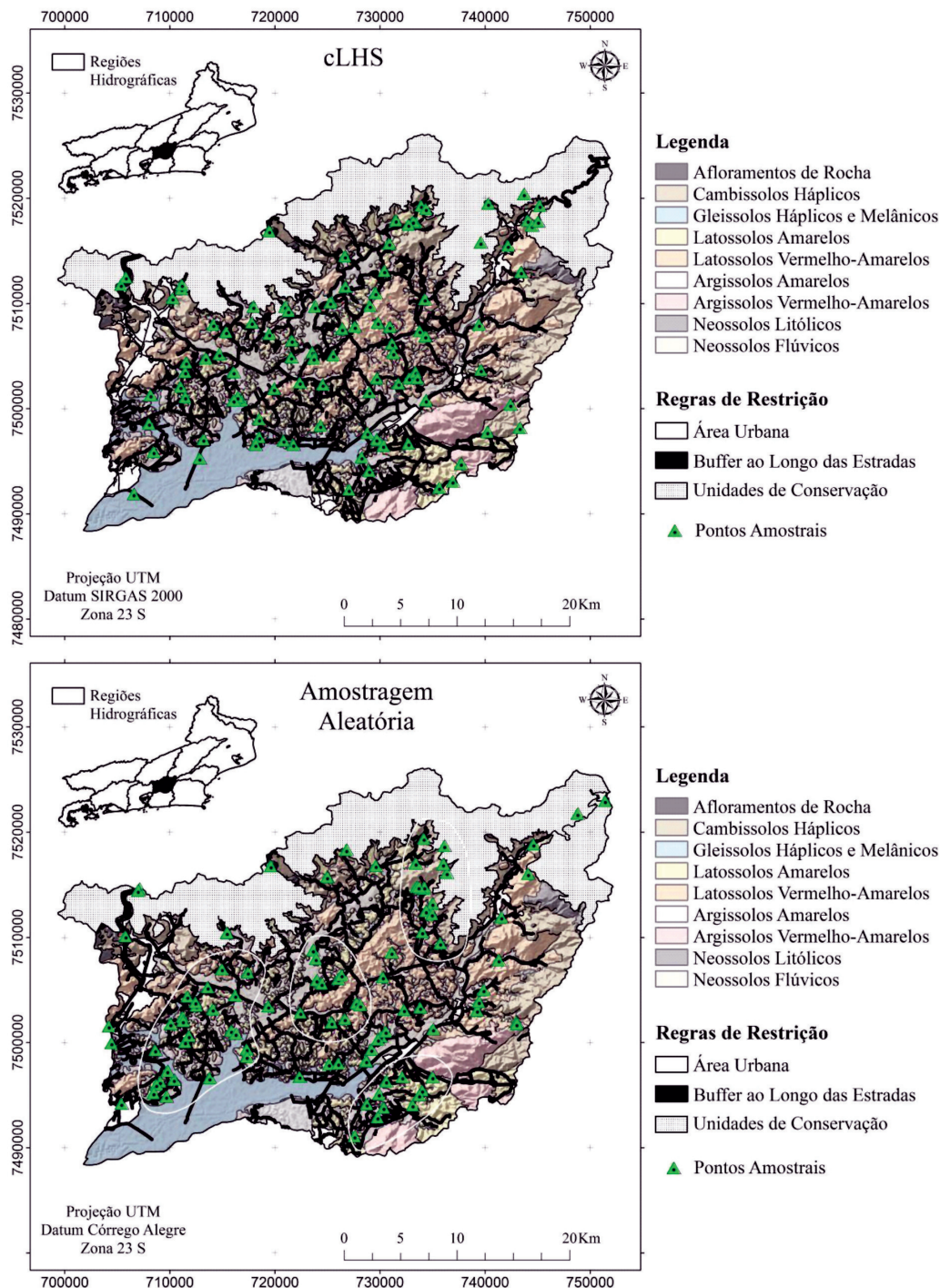


Figura 2. Distribuição espacial dos pontos amostrais determinados a partir do método cLHS e da amostragem aleatória na bacia dos Rios Guapi-Macacu.

(Quadro 1). Para a amostragem aleatória, observou-se diferença significativa para a distribuição dos dados na comparação com a bacia. Chagas et al. (2010) e Pinheiro (2012) em seus estudos verificaram estreita relação entre a curvatura do terreno e o grau de evolução dos solos. Na área da bacia, a curvatura influenciou diretamente a distribuição dos Latossolos e Cambissolos nas partes mais elevadas, com o primeiro ocupando as áreas convexas e o segundo as mais côncavas. Dessa maneira, a diferença observada entre a amostragem aleatória e a bacia poderia favorecer maior discrepância na amostragem desses solos, o que não foi verificado com a utilização do método cLHS. Sendo assim, o método cLHS mais uma vez foi mais eficiente para representar tanto a área da bacia quanto a área efetiva do que a amostragem aleatória.

Dentre as covariáveis ambientais contínuas utilizadas, a curvatura foi a que apresentou a maior redução do intervalo dos valores, tanto pelo método cLHS quanto pela amostragem aleatória. Essa redução, embora de certa maneira drástica, não teve nenhuma influência sobre a amostragem, em ambos os métodos, já que menos de 0,3 % da área da bacia e da área efetiva possuem curvatura <-2 ou >2 (Figura 1c). Resultados semelhantes foram obtidos por Thomas et al. (2012), para as variáveis perfil e plano de curvatura.

Na figura 1d, verifica-se que ambos os métodos utilizados preservaram bem a distribuição de frequência das classes de uso e cobertura do solo identificadas na área efetiva. A diferença observada com relação à bacia é em razão das regras de restrições utilizadas.

Com base nos resultados alcançados, foram escolhidos para a coleta das amostras os pontos selecionados pelo método cLHS, que representou a distribuição das covariáveis ambientais utilizadas melhor do que a amostragem aleatória, conforme a análise dos histogramas de frequência (Figura 1) e as análises estatísticas realizadas (Quadro 1). No quadro 2, é apresentada a relação completa das classes de solos identificadas e a relação dessas com as covariáveis ambientais utilizadas.

O elevado número de pontos amostrais identificados como Cambissolos Háplicos (18), Latossolos Amarelos (16) e Latossolos Vermelho-Amarelos (14) refletiu a ampla dominância dessas classes na bacia, conforme verificado por Embrapa (2003). O motivo de apenas um perfil de Argissolo Vermelho-Amarelo ter sido coletado pode estar relacionado com a não utilização da geologia como covariável ambiental, visto que esses são desenvolvidos de rochas alcalinas do Complexo Alcalino de Rio Bonito, diferentemente do material de origem dos Latossolos e Cambissolos. Sendo assim, este resultado pode ser interpretado como uma falha na escolha inicial da covariáveis.

Os Neossolos Litólicos ocorrem em condições de altitude, declividade e curvatura mais elevadas do que os Cambissolos (Quadro 2). Tendo em vista que a maior parte da área ocupada por essa classe ocorre dentro das Unidades de Conservação, o método cLHS selecionou apenas um ponto amostral para representá-la, respeitando a representatividade das condições ambientais de ocorrência na área efetiva. Os Nitossolos Háplicos, relacionados exclusivamente com alguns diques de diabásio/basalto, de ocorrência muito restrita na área da bacia (Silva & Cunha, 2001), não foram descritos no levantamento de reconhecimento de baixa intensidade dos solos do Estado do Rio de Janeiro (Embrapa, 2003); mesmo assim, o método cLHS selecionou três pontos amostrais para representar essa classe.

Os resultados apresentados no quadro 2 evidenciaram que os Argissolos Amarelos e os Planossolos Háplicos ocorrem em condições bastante similares de elevação, declividade e curvatura, normalmente em posição de terço inferior na transição entre os morrotes e a várzea, em relevo plano ou suave ondulado.

Já os Gleissolos Háplicos, Gleissolos Melânicos e Organossolos Háplicos foram descritos em áreas sujeitas a alagamento em terrenos muito mal-drenados da Planície Flúvio-Marinha e também apresentaram similaridade quanto às condições de elevação, declividade e curvatura do terreno (Quadro 2). Os Neossolos Flúvicos ocorrem nos entornos dos maiores

Quadro 1. Resultados das análises estatísticas realizadas

Covariável	Área	Teste F ⁽¹⁾		Teste T ⁽²⁾		Teste K-S ⁽³⁾	
				valor p			
		cLHS	AA	cLHS	AA	cLHS	AA
Elevação	Bacia	0,093	9,379e-08*	0,49	0,93	0,15	0,013*
	Efetiva	0,408	1,425e-05*	0,852	0,54	0,957	0,222
Declividade	Bacia	0,079	0,524	0,094	0,028*	0,286	0,0092*
	Efetiva	0,847	0,155	0,548	0,903	0,964	0,5739
Curvatura	Bacia	0,0699	0,1895	0,071	0,114	0,131	0,0214*
	Efetiva	0,3526	0,6537	0,647	0,858	0,870	0,4385

⁽¹⁾ Variância; ⁽²⁾ Média; ⁽³⁾ Distribuição; AA: amostragem aleatória; * estatisticamente diferentes a 95 %.

Quadro 2. Classes de solos identificadas nos 100 pontos selecionados pelo método cLHS

Classe de solo	Número de pontos	Covariável ambiental contínua ⁽¹⁾		
		Elevação	Declividade	Curvatura
		m	%	
Argissolos Amarelos	19	24,5	3,5	-0,03
Argissolos Vermelho-Amarelos	1	136,4	13,7	0,08
Cambissolos Háplicos	18	291,1	32,5	-0,16
Gleissolos Háplicos	12	15,4	2,0	-0,08
Gleissolos Melânicos	3	8,9	0,5	0,02
Latossolos Amarelos	16	57,5	21,2	0,09
Latossolos Vermelho-Amarelos	14	54,0	26,4	0,05
Neossolos Flúvicos	7	85,0	1,9	-0,11
Neossolos Litólicos	1	758,4	87,0	-1,40
Nitossolos Háplicos	3	37,8	4,3	-0,21
Organossolos Háplicos	3	15,4	0,8	0,02
Planossolos Háplicos	3	24,7	2,6	-0,01

⁽¹⁾ Média dos valores.

drenos, onde a dinâmica de alteração do fluxo, densidade de drenagem e vazão do rio ao longo dos tempos favoreceu a deposição de materiais de granulometrias distintas, originando as camadas/os horizontes. Nesse caso, a diferença mais acentuada para os Gleissolos e Organossolos foi verificada para a elevação, com os Neossolos Flúvicos ocorrendo em cotas mais elevadas (Quadro 2). Gleissolos Tiomórficos e Gleissolos Sálcos que ocorrem próximos à foz da bacia no entorno da Baía de Guanabara (Embrapa, 2003) não foram selecionados pelo método cLHS por causa da falta de estradas que excluiu essa porção da bacia da área efetiva utilizada.

O maior número de perfis de Argissolos Amarelos, Cambissolos Háplicos, Latossolos Amarelos, Latossolos Vermelho-Amarelos e Gleissolos Háplicos estão de acordo com a maior expressão dessas classes na área da bacia. A abordagem adotada se apresentou bastante satisfatória para as condições ambientais da bacia, onde a gênese dos solos é condicionada, principalmente pelas feições geomorfológicas.

CONCLUSÕES

1. A área efetiva obtida com a utilização das regras de restrição representou adequadamente as características da bacia em relação às covariáveis ambientais utilizadas, possibilitando aumentar a eficiência do trabalho, com menor custo e menor tempo, sem a perda da representatividade das condições ambientais.

2. O método cLHS, além de resultar em melhor distribuição geográfica, preservou a distribuição de frequência das covariáveis contínuas melhor do que a amostragem aleatória. Com relação à distribuição de

frequência da covariável categórica uso e cobertura do solo, os métodos demonstraram comportamento similar.

3. As amostras obtidas pelo método cLHS, estatisticamente, não apresentaram diferenças significativas para os testes estatísticos realizados com relação à bacia e com a área efetiva, enquanto a amostra aleatória evidenciou diferenças significativas para alguns desses parâmetros. Dessa maneira, o método cLHS pode ser considerado como apropriado para seleção de locais de amostragem para o mapeamento digital de solos, principalmente para áreas com condições similares as deste estudo, onde os recursos são limitados.

LITERATURA CITADA

- BRUNGARD, C.W. & BOETTINGER, J.L. Conditioned latin hypercube sampling: Optimal sample size for digital soil mapping of arid rangelands in Utah, USA. In: BOETTINGER, J.L.; HOWELL, D.W.; MOORE, A.C.; HARTEMINK, A.E. & KIENAST-BROWN, S., eds. Digital soil mapping: Bridging research, environmental application and operation. Dordrecht, Springer, 2010. p.67-75. (Progress in Soil Science, 2)
- BRUS, D.J. & HEUVELINK, G.B.M. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138:86-95, 2007.
- BRUS, D.J.; De GRUIJTER, J.J. & van GROENIGEN, J.W. Designing spatial coverage samples using the k-means clustering algorithm. In: LAGACHERIE, P.; McBRATNEY, A. & VOLTZ, M., eds. Digital soil mapping: An introductory perspective. Amsterdam, Elsevier, 2007. p.183-192. (Developments in Soil Science, 31)

- CARVALHO JÚNIOR, W.; CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.G.; BHERING, S.B. & FRANCELINO, M.R. Digital soilscape mapping of tropical hillslope areas by neural networks. *Sci. Agric.*, 68:691-696, 2011.
- CHAGAS, C.S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.G.R. & CARVALHO JÚNIOR, W. Atributos topográficos e dados do Landsat7 no mapeamento digital de solos com uso de redes neurais. *Pesq. Agropec. Bras.*, 45:497-507, 2010.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Levantamento de reconhecimento de baixa intensidade dos solos do Estado do Rio de Janeiro. Rio de Janeiro, 2003. 242p. (Boletim de Pesquisa e Desenvolvimento, 32)
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Manual de métodos de análise de solo. 2.ed. Rio de Janeiro, 1997. 212p. (Série Documentos, 1)
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Procedimentos normativos de levantamentos pedológicos. Brasília, 1995. 101p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Sistema brasileiro de classificação de solos. 2.ed. Rio de Janeiro, 2006. 306p.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Centro Nacional de Pesquisa de Solos. Uso e cobertura da terra na bacia hidrográfica do rio Guapi-Macacu. Rio de Janeiro, 2008. 31p. (Série Documentos, 105)
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE - ESRI. ArcGIS Desktop. Redlands, 2010. v.10.
- GALLANT, J.C. & WILSON, J.P. Primary topographic attributes. In: WILSON, J.P. & GALLANT, J.C., eds. *Terrain analysis: Principles and applications*. New York, John Wiley & Sons, 2000. p.51-85.
- GESSLER, P.E.; MOORE, I.D.; McKENZIE, N.J. & RYAN, P.J. Soil landscape modelling and spatial prediction of soil attributes. *Inter. J. Geogr. Inf. Syst.*, 9:421-432, 1995.
- HENGL, T.; ROSSITER, D.G. & STEIN, A. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Aust. J. Soil Res.*, 41:1403-1422, 2003.
- HENGL, T. Pedometric mapping: bridging the gaps between conventional and pedometric approaches. Enschede, Wageningen University, 2003. 200p. (Tese de Doutorado)
- HOWELL, D.; KIM, Y.; HAYDU-HOUDESHELL, C.; CLEMMER, P.; ALMARAZ, R. & BALLMER, R. Fitting soil property spatial distribution models in the Mojave Desert for digital soil mapping. In: LAGACHERIE, P.; McBRATNEY, A.B. & VOLTZ, M., eds. *Digital soil mapping: An introductory perspective*. Amsterdam, Elsevier, 2007. p.465-475. (Developments in Soil Science, 31)
- LESCH, S.M.; STRAUSS, D.J. & RHOADES, J.D. Spatial prediction of soil salinity using electromagnetic induction techniques 2: An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation. *Water Resour. Res.*, 31:387-398, 1995.
- McKAY, M.D.; BECKMAN, R.J. & CONOVER, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239-245, 1979.
- McKENZIE, N.J. & RYAN, P.J. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89:67-94, 1999.
- MINASNY, B. & McBRATNEY, A.B. Latin hypercube sampling as a tool for digital soil mapping. In: LAGACHERIE, P.; McBRATNEY, A.B. & VOLTZ, M., eds. *Digital soil mapping: An introductory perspective*. Amsterdam, Elsevier, 2007. p.153-165. (Developments in Soil Science, 31)
- MINASNY, B. & McBRATNEY, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.*, 32:1378-1388, 2006.
- PINHEIRO, H.S.K.; CHAGAS, C.S.; CARVALHO JÚNIOR, W. & ANJOS, L.H.C. Modelos de elevação para obtenção de atributos topográficos utilizados em mapeamento digital de solos. *Pesq. Agropec. Bras.*, 47:1384-1394, 2012.
- PINHEIRO, H.S.K. Mapeamento digital de solos por redes neurais artificiais da bacia hidrográfica do rio Guapi-Macacu, RJ. Seropédica, Universidade Federal Rural do Rio de Janeiro, 2012. 175p. (Dissertação de Mestrado)
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, 2013. Disponível em: <<http://www.r-project.org/isbn> 3- 900051-07-0>. Acesso em: 08 maio de 2013.
- ROECKER, S.M. & THOMPSON, J.A. Scale effects on terrain attribute calculation and their use as environmental covariates for digital soil mapping. In: BOETTINGER, J.L.; HOWELL, D.W.; MOORE, A.C.; HARTEMINK, A.E. & KIENAST-BROWN, S., eds. *Digital soil mapping: Bridging research, environmental application and operation*. Dordrecht, Springer, 2010. p.55-66. (Progress in Soil Science, 2)
- ROUDIER, P.; HEWITT, A.E. & BEAUDETTE, D.E. A conditioned latin hypercube sampling algorithm incorporating operational constraints. In: MINASNY, B.; MALONE, B.P. & McBRATNEY, A.B., eds. *Digital soil assessments and beyond*. London, CRC Press/Balkema, 2012. p.227-232.
- SANTOS, M.L.M.; DART, R.O. & BERBARA, R.L.L. Applying conditional Latin Hypercube (cLHS) for selecting soil sampling location for Digital Soil Mapping at Parque Estadual da Mata Seca, MG, Brazil. In: GLOBAL WORKSHOP ON DIGITAL SOIL MAPPING: Bridging research, production, and environmental application, 3., Logan, 2008. *Proceedings...* Logan, International Union of Soil Science (IUSS), 2008.
- SANTOS, R.D.; LEMOS, R.C.; SANTOS, H.G.; KER, J.C. & ANJOS, L.H.C. Manual de descrição e coleta de solo no campo. 5.ed. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, 2005. 100p.
- SILVA, L.C. & CUNHA, H.C.S. Geologia do Estado do Rio de Janeiro: Texto explicativo do mapa geológico do Estado do Rio de Janeiro. 2.ed. Brasília, CPRM, 2001. CD-ROM
- TEETOR, P. R Cookbook. Sebastopol, O'Reilly, 2011. 413p.

- THOMAS, M.; ODGERS, N.; RINGROSE-VOASE, A.; GREALISH, G.; GLOVER, M. & DOWLING, T. Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment. In: MINASNY, B.; MALONE, B.P. & McBRATNEY, A.B., eds. Digital soil assessments and beyond. London, CRC Press/Balkema, 2012. p.233-238.
- VÁŠÁT, R.; HEUVELINK, G.B.M. & BORUVKA, L. Sampling design optimization for multivariate soil mapping. *Geoderma*, 155:147-153, 2010.
- WEBSTER, R. & OLIVIER, M.A. Statistical methods in soil and land resource survey. Oxford, Oxford University Press, 1990. 316p.
- ZHU, A.X.; YANG, L.; LI, B.L.; QIN, C.Z.; ENGLISH, E.; BURT, J.E. & ZHOU, C.H. Purposive sampling for digital soil mapping for areas with limited data. In: HARTEMINK, A.E.; McBRATNEY, A.B. & MENDONÇA SANTOS, M.L., eds. Digital soil mapping with limited data. New York, Springer-Verlag, 2008. p.233-245.