Amostragem para mapeamento digital de solos: um turorial em R

Elias Mendes Costa, Marcos Bacis Ceddia, Felipe Nascimento dos Santos, Laiz de Oliveira Silva, Igor Prata Terra de Rezende, Douglath Alves Corrêa Fernandes

# Introdução

## Amostragem

A amostragem diz respeito à seleção de um subconjunto de indivíduos de uma população para estimar características de toda a população (Wang et al., 2012).

## Amostragem no mapeamento convencional de solos

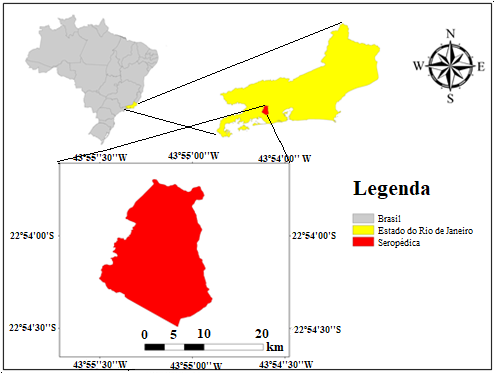
No caso do mapeamento convencional de solos, o método de prospecção e a frequência de amostragem dependem do nível de detalhe do levantamento e do objetivo do levantamento (IBGE, 2015). Por exemplo, na terceira edição do Manual Técnico de Pedologia (IBGE, 2015) o método de prospecção indicado para levantamento exploratório é o de extrapolação, generalizações, correlações e poucas observações de campo, com uma frequência de amostragem de um perfil completo por classe de solo predominante na associação. Para o levantamento semidetalhado a densidade de observações e a frequência de amostragem são calculadas em função da heterogeneidade da área e da facilidade de correlação de solos e a superfície geomorfológicas. No entanto, é recomendada uma média de 0,002 a 0,20 observações por hectare e um perfil completo e um complementar por classe de solo componente da unidade de mapeamento simples ou associação. Porém para o MDS esse esquema de amostragem não se aplica, pois não fornece estimativa estatística representativa. No MDS é necessário o uso de estratégias de amostragem estatisticamente robustas para diminuir a subjetividade e consequentemente os possíveis erros na predição (Minasny and McBratney, 2007)

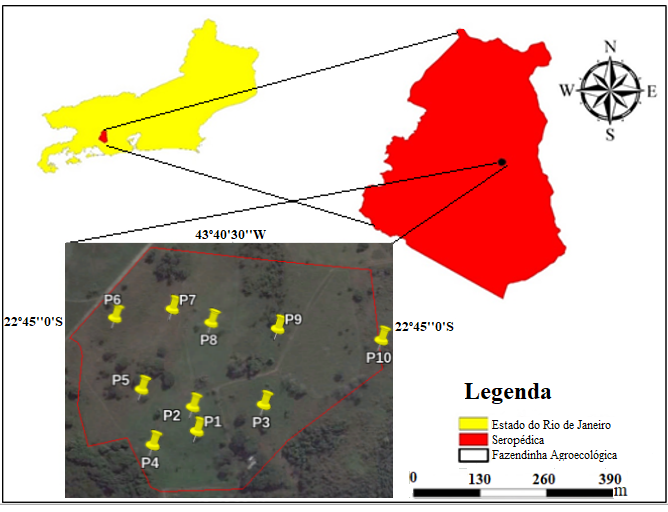
## Amostragem no mapeamento digital de solos

No caso do mapeamento digital de solos (MDS) a amostragem é um fator crucial, pois pode contribuir para elevar ou diminuir os custos do levantamento de solos. Ainda é considera uma das etapas mais importantes para o MDS (Carvalho Junior et al., 2014). Para o MDS, para uma otimização da calibração de modelos matemáticos as amostras coletadas devem representar ao máximo a distribuição espacial do que está sendo modelado (atributo ou classe de solo), por isso técnicas como o Hipercubo Latino Condicionado (cLHS) (Minasny and McBratney, 2006) são frequentemente usados nessa etapa e para essa finalidade. Esse tipo de amostragem é conhecido como amostragem não probabilística e será abordado com mais detalhes nos exemplos. Para validação dos resultados das predições espaciais no MDS o tipo de abordagem recomendado é a amostragem probabilística (Brus et al., 2011), esse tipo de amostragem busca a imparcialidade e diminuição de viés para validação dos produtos do mapeamento. Atenção especial deve ser tomada nessa etapa, pois nem sempre é possível uma amostragem completamente probabilística (aleatória). Há então a necessidade de encontrar a condição perfeita do que é factível e o ideal. Exemplos de amostragem probabilística serão abordados com mais detalhe nos exemplos.  
Parte do material aqui desenvolvido foi extraído e adaptado do livro em desenvolvimento de Alessandro Samuel-Rosa, [pedometria feita simples](http://samuel-rosa.github.io/pedometria-feita-simples/principios-da-amostragem-espacial.html). O capítulo 7 do livro trata de princípios da amostragem espacial. Também foram extraídas informações a respeito de amostragem para validação de mapas digitais de solos (Brus et al., 2011) e amostragem para mapeamento de solos (Brus, 2019). Assim objetivo desse tutorial e compilar as informações dos trabalhos mencionados para simular difentes estratégias amostrais para dar suporte aos levantamentos de solos com foco em MDS e avaliação dos produtos gerados.

# Área experimental e base dados

## Área experimental

Ao longo desse tutorial sobre amostragem, você verá exemplos de cada tipo de amostragem implementados no R usando o conjunto de dados de um área experimental localizada no município de Seropédica, estado do Rio de Janeiro, Brasil, nas coordenadas geográficas centrais aproximadas 22º54’00” S e 43º55’00” W (Figura 1). 

Especificamente, o trabalho foi desenvolvido na Fazendinha Agroecológica km 47 (Sistema Integrado de Produção Agroecológica - SIPA), associada à Embrapa Agrobiologia, Empresa de Pesquisa Agropecuária do Estado do Rio de Janeiro (Pesagro-Rio) e Universidade Federal Rural do Rio de Janeiro, nas coordenadas geográficas centrais aproximadas 22º45’0” S e 43º40’30” W (Figura 2). 

O clima do município caracteriza-se por ser do tipo Aw (Tropical com inverno seco), segundo a classificação de Köppen. Seropédica tem uma temperatura média de 23,5 °C e uma pluviosidade média anual de 1354 mm. Os maiores valores de precipitação são verificados nos meses de novembro a janeiro e os menores no período entre maio e agosto (Oliveira-Júnior et al., 2014). A área dos exemplos encontra-se ocupada por pastagem e caracteriza-se por apresentar relevo suave-ondulado a ondulado, altitude entre 17 a 50 m sobre o nível do mar, e ausência de rochosidade e pedregosidade em superfície (Nascimento, 2019).

Os Solos da área experimental são caraterizados por apresentar, no geral baixa fertilidade e elevado grau de desonvilvimento com fatores e processos de formação específicos e dominates para cada ordem e posição da paisagem. Para o trabalho de onde os dados derivam (Nascimento, 2019), foram abertas 10 trincheiras em diferentes pontos da paisagem, onde foram descritos e classificados perfis do solo, sendo uma trincheira no topo (P10), duas na posição de terço médio (P3 e P9), uma no terço inferior (P8), três no sopé da vertente (P2, P5 e P7) e três na área de baixada (P1, P4 e P6). Os solos froam classificados respectivamente como Argissolos Vermelhos P10, P9 e P3; Argissolo Amarelo, P8; Planossolos Háplicos (arênicos e espessarenicos), P7,P2,P5; Planossolos Háplicos (gleissólicos), P4, P6 e Cambissolo Hálico (com evidente processos de gleização) P1 (Figura 2).

## Base de dados (Covariáveis ambientais)

As covariáveis ambientais para modelar as propriedades do solo foram derivadas de duas fontes de dados: modelo digital de elevação, dados de sensoriamento remoto (imagem orbital), uma imagem RapidEye. Elas foram escolhidos para descrever os principais fatores formadores do solo, de acordo com a abordagem scorpan (McBratney et al., 2003).

*Modelo Digital de Elevação (MDE)*: O MDE utilizado, com resolução espacial de 2 m, foi gerado a partir das curvas de nível e pontos cotados coeletados com gps geodésico. Do MDE foram derivados 12 atributos de terreno com pontencial uso para modelagem no MDS.

*Imagem de satélite*: foram utilizadas duas cenas do sensor RapidEyeambas de 2011. As duas cenas foram usadas para cobrir toda a área de estudo e eles têm uma 12 resolução radiométrica de bits, resolução espacial de 6,5m, e ortorretificadas para resolução espacial de 5m (RapidEye, 2012). Das cenas foram derivados os indices de vegetação (NDVI, inglês) e índice de vegetação ajustado ao solo (SAVI, inglês).

## Áreas ambientalmente homogêneas (Estratos)

Para fazer o exerício foram criadas áreas ambientalmente similares (homogêneas), essa a bordagem é indicada em casos onde há por exemplo o interesse de fazer a amostragem po estratos. Nesse execício optou-se por criar 10 estratos usando as variáveis, que julgamos serem as mais representativas para discretizar os ambientes. Foram elas elevação, declividade e SAVI. Para fazer o agrupamento da áreas de maior similaridade (criação de clusters) foi utilizado o algorítimo de classificação não supervisionada *K-Means Clustering* que visa particionar os pontos em k grupos de forma que a soma dos quadrados dos pontos aos centros de cluster designados seja minimizada. No mínimo, todos os centros do cluster estão na média dos seus conjuntos de Voronoi (o conjunto de pontos de dados que estão mais próximos do centro do cluster).

## Carregar os pacotes  
library(sp)  
library(raster)  
library(sampling)  
library(maptools)  
library(rgdal)  
library(clhs)  
  
# Carregando a base de dados da fazendinha  
dados <- read.csv("../data/dataset.csv"); dados=dados[2:54]  
grid <- read.csv("../data/grid.csv"); grid=grid[2:23]  
  
# criando um cluster de unidades homogêneas usando os dados das covariáveis ambientais  
cluster <- kmeans(grid[, c(1,2,20)], 10)  
cluster = data.frame(cluster$cluster)  
grid=cbind(grid,cluster)  
  
## Foi necessário transformar os estratos para fator   
##porque estava dando erro quando calculava o número  
## de amostras por estrato   
  
# Função para transformar uma variável em fator  
grid$cluster.cluster=as.factor(grid$cluster.cluster)   
  
#cluster=data.frame(grid[,c("X","Y")],results=cluster$cluster)  
#coordinates(cluster)= ~ X+Y; gridded(cluster)=T; cluster=raster(cluster)  
#plot(cluster)  
  
# Transformar o dataframe do grid em spatialdataframe  
sp::gridded(grid) <- ~ X + Y  
sp::proj4string(grid) <- sp::CRS("+proj=utm +zone=23 +south +ellps=GRS80 +  
 towgs84=0,0,0,0,0,0,0 +units=m +no\_defs")   
  
# Definindo o sistema de projção  
crsSIRGAS2000=CRS("+proj=utm +zone=23 +south +ellps=GRS80 +  
 towgs84=0,0,0,0,0,0,0 +units=m +no\_defs")   
  
#carregando arquivos shapfile  
## Estradas  
estrada <- maptools::readShapeSpatial("../shape/estrada.shp",   
 proj4string=crsSIRGAS2000,verbose=TRUE)

## Shapefile type: PolyLine, (3), # of Shapes: 1  
## Shapefile type: PolyLine, (3), # of Shapes: 1

## Perfis completos  
perfis <- maptools::readShapeSpatial("../scripts/pontos.shp",   
 proj4string=crsSIRGAS2000,verbose=TRUE)

## Shapefile type: Point, (1), # of Shapes: 71  
## Shapefile type: Point, (1), # of Shapes: 71

# Amostragem para mapeamento de solos

Não exite uma um único tipo de estratégia amostral para qualquer trabalho de MDS, cada caso precisa ser estudado de forma particular e no geral a melhor estratégia depnede so objetivo do projeto, do recurso financeiro e humano disponível e dificuldades operacionais encontradas na área que está sendo estuda. Do ponto de vista de dificuldades operacinais tem-se alguns variantes como limitação de de acesso, por se tratar de áreas com acesso limitado, nesses casos há abordagens de amostragem específica como mostra Cambule et al. (2013) onde se tem áreas de maior acessibilidade e que sejam representativas da área total de estudo. Ainda pode haver limitação por permissão de acesso, sobretudo em áreas particulares. Quanto aos tipos de amostragem elas são divididas basicamente em dois grupos ou tipos fundamentais: *Amostragem probabilística* e *não probabilística*

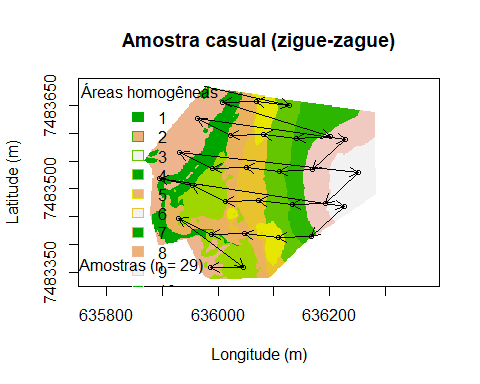
## Amostragem não probabilística

A amostragem não probabilística é uma técnica de amostragem em que as amostras são reunidas em um processo que não dá a todos os indivíduos, no caso de solos perfil de solo *pedon* ou ponto amsotral, da população chances iguais de serem selecionados. A escolha dos locais de amostragem depende da definição de um critério a ser atendido, normalmente em MDS a caracterização da máxima variação de solos e suas propriedades representados pelas covariáveis ambientais que simulam os fatores de formação do solo (McBratney et al., 2003). A amostragem não-probabilística costuma ser dividida em três categorias: casual, conveniente, e intencional.

### Amostragem casual

O texto vem aqui o texto vem aqui o texto vem aqui

main <- "Amostra casual (zigue-zague)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(3),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
set.seed(2000)  
pts <- sp::spsample(grid, n = 30, type = "nonaligned")  
points(pts, pch = 21, cex = 0.75)  
arrows(  
 grid@coords[1, 1], grid@coords[1, 2], pts@coords[length(pts), 1],  
 pts@coords[length(pts), 2], length = 0.1)  
for (i in 2:length(pts) - 1) {  
 arrows(  
 x1 = pts@coords[i, 1], y1 = pts@coords[i, 2], x0 = pts@coords[i + 1, 1],   
 y0 = pts@coords[i + 1, 2], length = 0.1)  
}  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



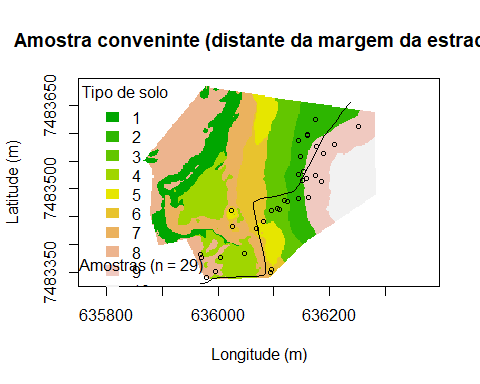
### Amostragem conveniente

O texto vem aqui o texto vem aqui o texto vem aqui Aqui foram selecionadas amostras com distância máxima da estrada de 50 m

main <- "Amostra conveninte (distante da margem da estrada)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Tipo de solo", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
plot(estrada, add=TRUE)  
t=buffer(estrada, width=50)

## Loading required namespace: rgeos

r <- raster(grid)   
lr <- mask(r, t)  
lr=as(lr, "SpatialPixelsDataFrame")  
points(spsample(lr, n = 30, "random"), pch = 21, cex = 0.75)  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")

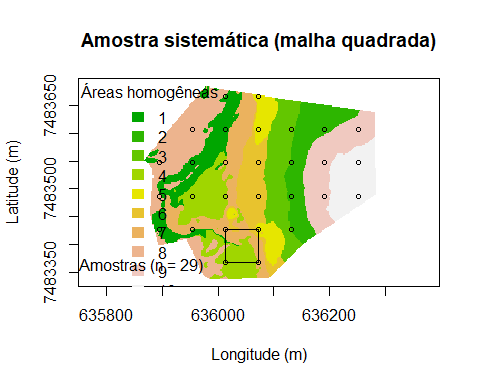


### Amostragem intencional

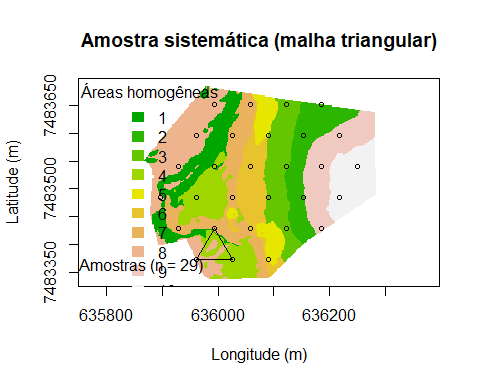
O texto vem aqui o texto vem aqui o texto vem aqui

#### Amostragem intencional sistemática

main <- "Amostra sistemática (malha quadrada)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
pts <- sp::spsample(grid, n = 30, type = "regular", offset = c(0.5, 0.5))  
points(pts, pch = 21, cex = 0.75)  
polygon(pts@coords[c(1, 2, 6, 5, 1), ])  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")

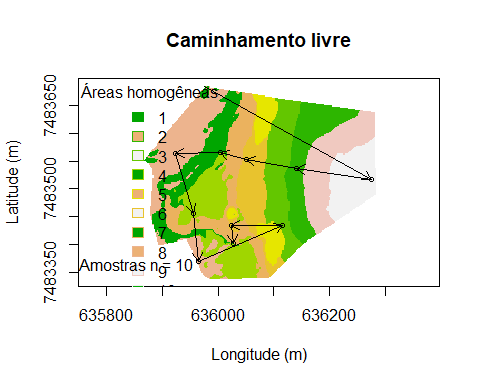


main <- "Amostra sistemática (malha triangular)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
pts <- sp::spsample(grid, n = 34, type = "hexagonal", offset = c(0.5, 0.5))  
points(pts, pch = 21, cex = 0.75)  
polygon(pts@coords[c(1, 2, 5, 1), ])  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



#### Amostragem intencional Caminhamento livre

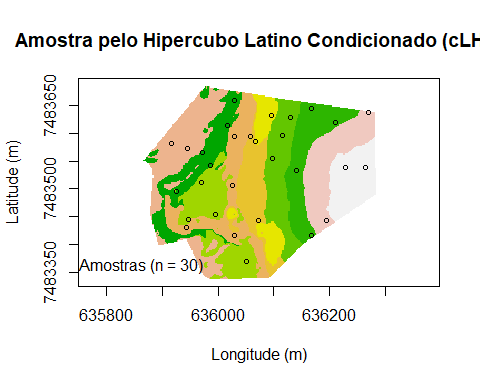
main <- "Caminhamento livre"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(3),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
points(perfis, pch = 21, cex = 0.75)  
arrows(  
 grid@coords[1, 1], grid@coords[1, 2], perfis@coords[length(perfis), 1],   
 perfis@coords[length(perfis), 2],  
 length = 0.1)  
for (i in 2:length(perfis) - 1) {  
 arrows(  
 x1 = perfis@coords[i, 1], y1 = perfis@coords[i, 2], x0 = perfis@coords[i + 1, 1],   
 y0 = perfis@coords[i + 1, 2], length = 0.1)  
}  
leg <- paste("Amostras n = 10", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



### Hipercubo Latino Condicionado (cLHS)

O cLHS é um algoritmo de busca baseado em regras heurísticas combinadas com um cronograma de recozimento. Para amostragem de uma área informações prévias representativas da área de estudo são fornecidas ao algoritmo como dados auxiliares (covariáveis ambientais). Ele fornece uma cobertura completa da faixa de cada covariável estratificando a distribuição marginal, ou seja, usa as informações fornecidas para produzir uma estratificação de amostragem otimizada (Minasny and McBratney, 2006). Variações do cLHS para otimização de locais de amostragem podem ser encontradas em Roudier et al. (2012), Mulder et al. (2013), Carvalho Junior et al. (2014), Clifford et al. (2014), Kidd et al. (2015), e Stumpf et al. (2016). No caso desse exemplo usado no cLHS como variáveis auxiliares a elevação, declividade e SAVI, ou seja as mesmas que foram escolhidas para gerar as áreas ambientalmente homogêneas (os estratos).

dem=raster(grid[1])  
slope=raster(grid[2])  
savi=raster(grid[20])  
grid1=raster::addLayer(dem, slope, savi)  
s <- rasterToPoints(grid1, spatial=TRUE)  
set.seed(2001)  
pts <- clhs(s, size = 30, iter = 1000, progress = FALSE)   
main <- "Amostra pelo Hipercubo Latino Condicionado (cLHS)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
points(pts, pch = 21, cex = 0.75)  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



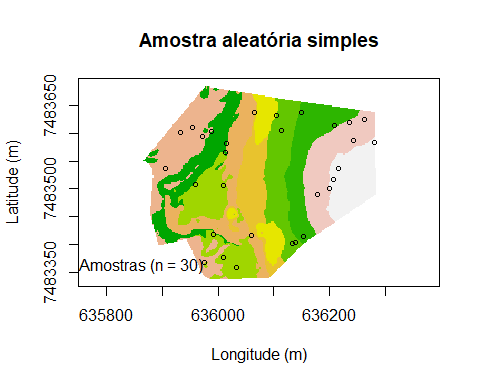
## Amostragem probabilística

A amostragem probabilistica é uma técnica de amostragem em que a amostra de uma população maior, no caso uma coleção de perfis de solo ou *pedon* de toda área, é escolhida usando um método baseado na teoria da probabilidade. Para que um perfil (ou ponto amostral) seja considerado uma amostra probabilística, ele deve ser selecionado usando uma seleção aleatória. Nesse tipo de amostragem todo e qualquer local possui alguma chance de ser amostrado, mesmo que alguns tenham maior chance do que outros. No caso da modelagem espacial do solo, a amostragem probabilística costuma ser usada para a validação das predições espaciais (Brus et al., 2011). Entretanto, ela também pode ser usada para obter observações para a calibração dos modelos preditivos ou serem usadas para validação e calibração em métodos de calçibração cruzada como o *leave-one-out* (LOO-CV) (Brus et al., 2011).

### Amostragem aleatória simples

O texto vem aqui o texto vem aqui o texto vem aqui

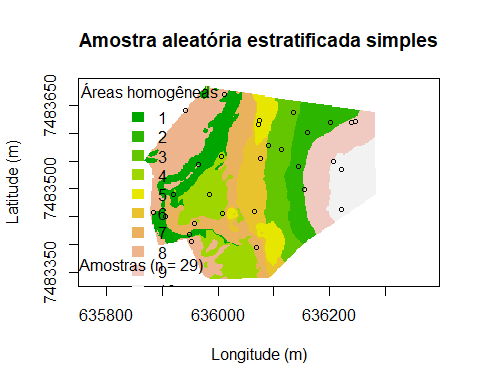
main <- "Amostra aleatória simples"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
set.seed(2001)  
pts <- grid[sampling::srswr(30, length(grid)) == 1, ]  
set.seed(2001)  
pts@coords <-   
 pts@coords + matrix(runif(prod(dim(pts@coords)), min = -0.5, max = 0.5), ncol = 2) \*   
 grid@grid@cellsize  
points(pts, pch = 21, cex = 0.75)  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



### Amostragem aleatória estratificada simples

O texto vem aqui o texto vem aqui o texto vem aqui

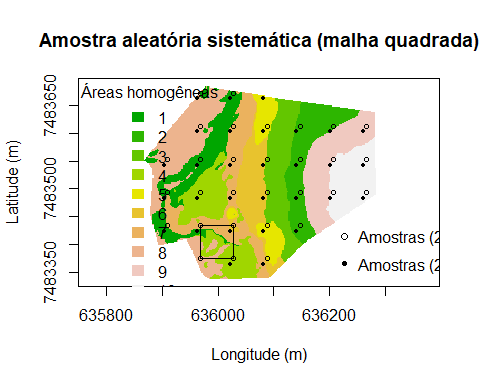
main <- "Amostra aleatória estratificada simples"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend( "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
n <- round(30 \* summary(grid$cluster.cluster) / length(grid))  
set.seed(2001)  
pts <- sampling::strata(  
 grid[order(grid$cluster.cluster), ], stratanames = "cluster.cluster", size = n,   
 method = "srswr")$ID\_unit  
pts <- grid[order(grid$cluster.cluster), ][pts, ]  
set.seed(2001)  
pts@coords <- pts@coords +   
 matrix(runif(prod(dim(pts@coords)), min = -0.5, max = 0.5), ncol = 2) \*   
 grid@grid@cellsize  
points(pts, pch = 21, cex = 0.75)  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



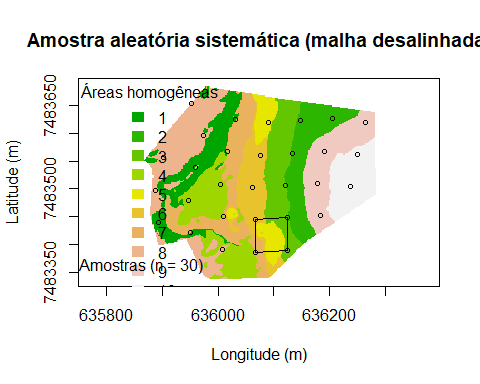
### Amostragem aleatória sistemática

O texto vem aqui o texto vem aqui o texto vem aqui

main <- "Amostra aleatória sistemática (malha quadrada)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
set.seed(2001)  
pts <- sp::spsample(grid, n = 30, type = "regular")  
points(pts, pch = 21, cex = 0.75)  
polygon(pts@coords[c(1, 5, 6, 2, 1), ])  
set.seed(1984)  
points(sp::spsample(grid, n = 30, type = "regular"), pch = 20, cex = 0.75)  
legend(636200, 7483450, legend = "Amostras (2019)", pch = 21, bty = "n")  
legend(636200, 7483400, legend = "Amostras (2020)", pch = 20, bty = "n")



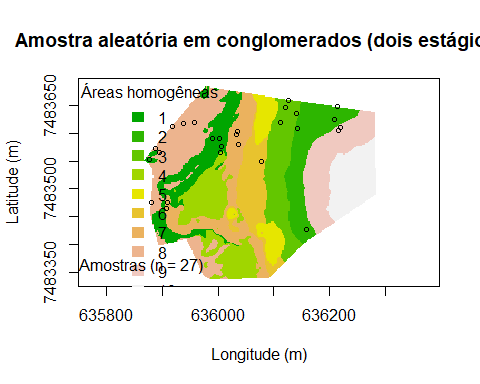
main <- "Amostra aleatória sistemática (malha desalinhada)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
set.seed(2001)  
pts <- sp::spsample(grid, n = 31, type = "nonaligned")  
points(pts, pch = 21, cex = 0.75)  
polygon(pts@coords[c(2, 7, 8, 3, 2), ])  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



### Amostragem aleatória em conglomerados

O texto vem aqui o texto vem aqui o texto vem aqui

main <- "Amostra aleatória em conglomerados (dois estágios)"  
plot(  
 grid@coords, type = "n", asp = 1, main = main,  
 xlab = "Longitude (m)", ylab = "Latitude (m)")  
image(  
 grid, "cluster.cluster", col = terrain.colors(10), axes = TRUE, add = TRUE)  
legend(  
 "topleft", title = "Áreas homogêneas", fill = terrain.colors(10),  
 legend = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),  
 border = terrain.colors(10), bty = "n")  
set.seed(2002)  
pts <- sp::spsample(grid, n = 33, type = "clustered", nclusters = 10)  
points(pts, pch = 21, cex = 0.75)  
leg <- paste("Amostras (n = ", length(pts), ")", sep = "")  
legend(635700, 7483400, legend = leg, pch = 21, bty = "n")



# References

Brus, D.J., 2019. Sampling for digital soil mapping: A tutorial supported by R scripts. Geoderma 338, 464–480. <https://doi.org/10.1016/j.geoderma.2018.07.036>

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62, 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364.x>

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., 2013. A methodology for digital soil mapping in poorly-accessible areas. Geoderma 192, 341–353. <https://doi.org/10.1016/j.geoderma.2012.08.020>

Carvalho Junior, W., Lagacherie, P., Chagas, C. da S., Calderano Filho, B., Bhering, S.B., 2014. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. Geoderma 232-234, 479–486. <https://doi.org/10.1016/j.geoderma.2014.06.007>

Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. Computers and Geosciences 67, 62–68. <https://doi.org/10.1016/j.cageo.2014.03.005>

IBGE, 2015. Manual Técnico de Pedologia, 3 Edição. ed. Rio de Janeiro.

Kidd, D., Malone, B., McBratney, A.B., Minasny, B., Webb, M., 2015. Operational sampling challenges to digital soilmapping in Tasmania, Australia. Geoderma Regional 4, 1–10.

McBratney, A.B., Mendonça-Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. <https://doi.org/10.1016/S0016-7061(03)00223-4>

Minasny, B., McBratney, A., 2007. Latin hypercube sampling as tool for digital soil mapping. Developments in Soil Science 31, 153–606. <https://doi.org/10.1016/S0166-2481(06)31012-4>

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers and Geosciences 32, 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>

Mulder, V.L., Bruin, S. de, Schaepman, M.E., 2013. Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. International Journal of Applied Earth Observations and Geoinformation 21, 301–310. <https://doi.org/10.1016/j.jag.2012.07.004>

Nascimento, C.W.R., 2019. Carlos wagner rodrigues do nascimento 2019 (PhD thesis).

Oliveira-Júnior, J.F., Delgado, R.C., Gois, G., Lannes, A., Dias, F.O., Souza, J.C., Souza, M., 2014. Análise da precipitação e sua relação com sistemas meteorológicos em seropédica, Rio de Janeiro. Floresta e Ambiente 21, 140–149. <https://doi.org/10.4322/floram.2014.030>

RapidEye, 2012. RapidEye Mosaic™ Product Specifications.

Roudier, P., Hewitt, A.E., Beaudette, D.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. Digital Soil Assessments and Beyond 227–231. <https://doi.org/10.1201/b12728-46>

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., Scholten, T., 2016. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. Journal of Plant Nutrition and Soil Science 000, 1–11. <https://doi.org/10.1002/jpln.201500313>

Wang, J.F., Stein, A., Gao, B.B., Ge, Y., 2012. A review of spatial sampling. Spatial Statistics 2, 1–14. <https://doi.org/10.1016/j.spasta.2012.08.001>