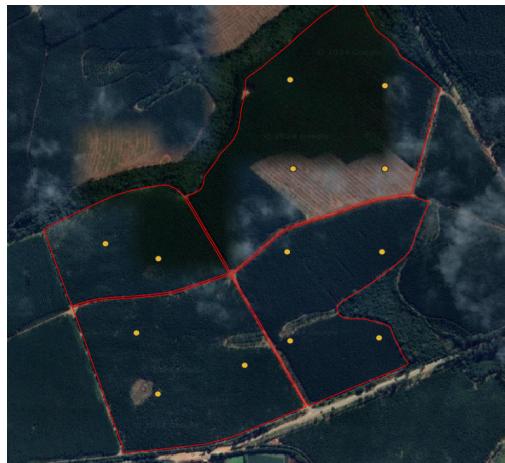


Descrição das etapas de processamento de dados de inventário realizado com o auxílio da tecnologia LiDAR

Otávio Magalhães Silva Souza



Piracicaba, SP – Data de Emissão: 26 de julho de 2024

Contents

1 Pacotes utilizados no R (colocar breve descrição - já tem uma descriçaozinha no R passado em aula)	3
1.1 Tidyverse (https://livro.curso-r.com/4-2-tidyverse.html)	3
1.2 Sf	3
1.3 Tidyterra	4
1.4 Terra	4
1.5 Stars	4
1.6 Tools	4
1.7 RColorBrewer	4
1.8 Progress	4
1.9 Reshape2	4
1.10 Mapview	4
1.11 LidR	4
1.12 RCSF	4
1.13 Future	4
2 Descrição da área	5
3 Grid e parcelas já inventariadas	6
4 Conceitos e definições da Dupla Amostragem	7
5 Fluxograma e etapas Dupla amostragem	7
6 Fluxograma e etapas Tripla amostragem	23

1 Pacotes utilizados no R (colocar breve descrição - já tem uma descriçaozinha no R passado em aula)

1.1 Tidyverse (<https://livro.curso-r.com/4-2-tidyverse.html>)

O Tidyverse é um pacote guarda-chuva e contém diversas funções úteis para garantir o dinamismo no script, visualização, processamento e análise dos dados, modelagem etc.

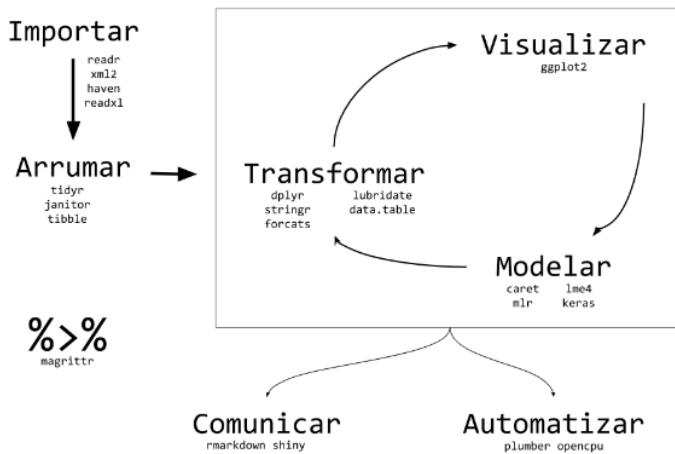


Figure 1: Tidyverse

1.2 Sf

Pacote utilizado para manipulação de objetos do mundo real. Descreve a forma com que esses objetos podem ser armazenados e importados e quais operações geométricas podem ser definidas por eles.

1.3 Tidyterra

1.4 Terra

1.5 Stars

1.6 Tools

1.7 RColorBrewer

1.8 Progress

1.9 Reshape2

1.10 Mapview

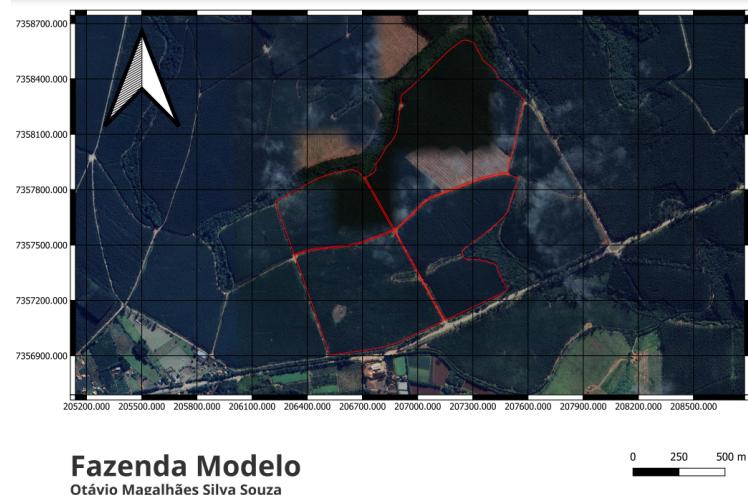
1.11 LidR

1.12 RCSF

1.13 Future

2 Descrição da área

A área a ser estudada como “Fazenda Modelo” localiza-se no município de São Miguel Arcanjo (SP), pode ser identificada pelas coordenadas (-23.86707°, -47.87772°) e possui 129,784 ha, que dividem-se em 4 subtalhões: 301a (18,933 ha), 301d (34,468 ha), 302a (47,602 ha) e 302c (28,781 ha).



Fazenda Modelo
Otávio Magalhães Silva Souza

Figure 2: Mapa da propriedade

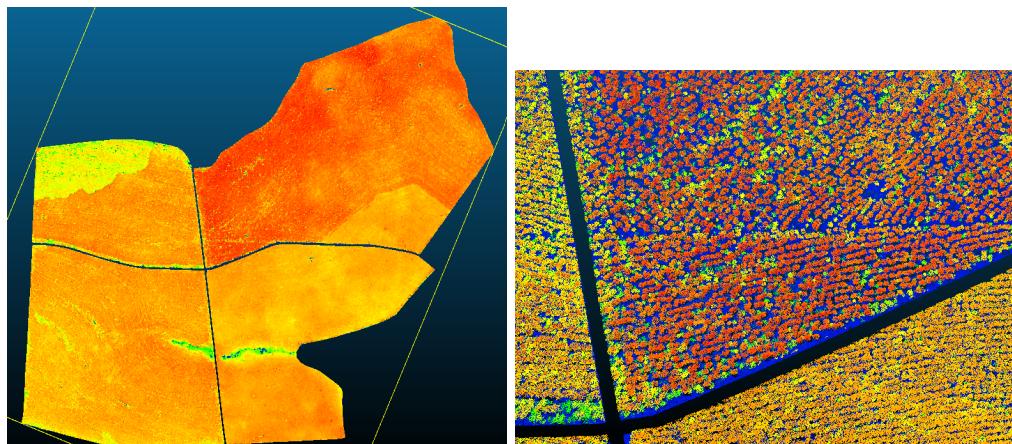


Figure 3: Nuvens LiDAR normalizadas

3 Grid e parcelas já inventariadas

A região foi dividida em 3454 parcelas, onde 2960 delas possuem 400m², enquanto as outras são menores por estarem na borda e abrangerem áreas além da área de interesse. Além disso, 13 das parcelas possuem dados de inventário florestal e podem ser identificadas pelos seguintes Id's: 993, 1526, 1770, 1881, 3165, 3628, 3660, 3730, 5052, 5091, 5106 e 5122.



Figure 4: Parcelas com dados de inventário

4 Conceitos e definições da Dupla Amostragem

A dupla amostragem é composta por duas fases (s1 e s2) - s1: compreende uma gama de variáveis explanatórias para cada ponto pertencente a s1. As variáveis explanatórias derivam de informações auxiliares disponíveis em grande quantidade ao longo da área florestal;

- s2: constitui o inventário terrestre feito num número limitado de subamostras, onde todas pertencem à s1 e fornecem o valor das variáveis de interesse, ex. densidade local.

Exemplo prático e definições (<https://www.ipef.br/publicacoes/scientia/nr108/cap09.pdf>)

- O erro amostral (E%) é calculado a partir do desvio padrão S^2 , que representa a variação de uma série de médias retiradas da população
- A dupla amostragem se divide em duas fases. Na primeira fase há uma maior intensidade amostral e se é medida a variável auxiliar, que deve ser de mensuração facilitada. Na segunda fase a intensidade amostral é menor e se mede a variável de interesse e a variável auxiliar, pois será calculado um estimador de regressão entre as duas. (existe uma diferença entre estimadores de regressão e de razão - estudar)
- A alta correlação entre métricas LiDAR e parâmetros biofísicos da floresta justificam a adoção da tecnologia na primeira fase da DA, além de reduzir a intensidade amostra (custo e trabalho) - Primeira fase: lançamento das parcelas e sobrevôo com o LiDAR
- Etapas (2a fase): mensuração de variáveis das árvores (ex. DAP), medição de algumas árvores para gerar modelo hipsométrico, que será usado para descobrir a altura das outras árvores, cálculo do volume também por modelo e extração do volume da parcela.
- O estimador de regressão (estudar) do VTCC deve relacionar a(s) variável(es) auxiliar(es) e também compreende a outros índices (constantes para cada região)
- O desvio padrão, intervalo de confiança etc são feitos depois



Figure 5: Fluxograma DA e estimador de regressão para VTCC

5 Fluxograma e etapas Dupla amostragem

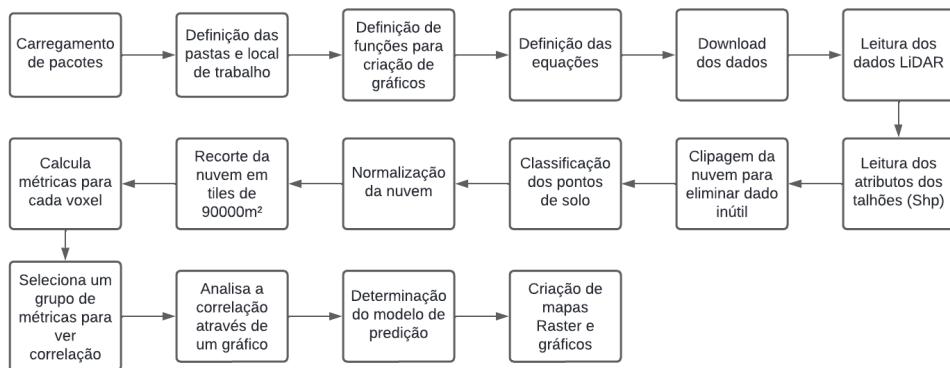


Figure 6: Fluxograma das etapas de processamento de dados LiDAR para fins de inventário florestal

1. Carregamento dos pacotes

- i. Diversos são os pacotes carregados. Os nomes e a utilidade de cada um estão descritos na primeira seção do documento.

2. Definição das pastas e local de trabalho

- GitHub
 - i. C: - pasta raiz
 - ii. GitRepo - diretório em que estão agrupados os arquivos a serem upados no GitHub
 - iii. PRJ_FAZENDAMODELO - pasta do projeto
 - iv. RMD - código e arquivos utilizados na redação do presente documento
 - v. RESULTADOS - plot da matriz de correlação
 - vi. BATCHR - arquivo tipo R com os scripts utilizados no pré e pós-processamento dos dados
 - vii. SAIDASSIG - arquivos gerados no QGIS
 - viii. SHAPES - arquivos de entrada para uso no SIG
- LiDAR
 - i. C: - pasta raiz
 - ii. LiDAR - agrupa todos as nuvens de pontos utilizadas no script
 - iii. PRJ_FAZENDAMODELO - pasta do projeto
 - iv. NUVENS - onde se localizam as nuvens de pontos
 - v. A13 - reúne as nuvens do ano de 2013
 - vi. TALHOES - nuvem segregada por talhão
 - vii. NoNORM - nuvens com solo classificado
 - viii. SiNORM - nuvens com solo classificado e normalizadas
 - ix. RSTR_qua - raster apresentando a estimativa da variável de interesse para cada talhão

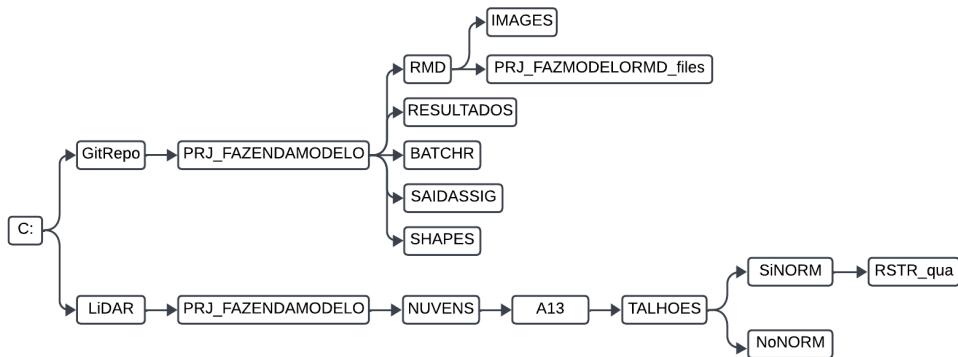


Figure 7: Organização dos diretórios

3. Definição das funções para criação dos gráficos

4. Definição das equações (estudar quais são)

5. Download e leitura dos dados LiDAR
 - i. Ao todo foram baixadas 6 nuvens de pontos LiDAR, que antes do processamento encontravam-se da seguinte maneira:

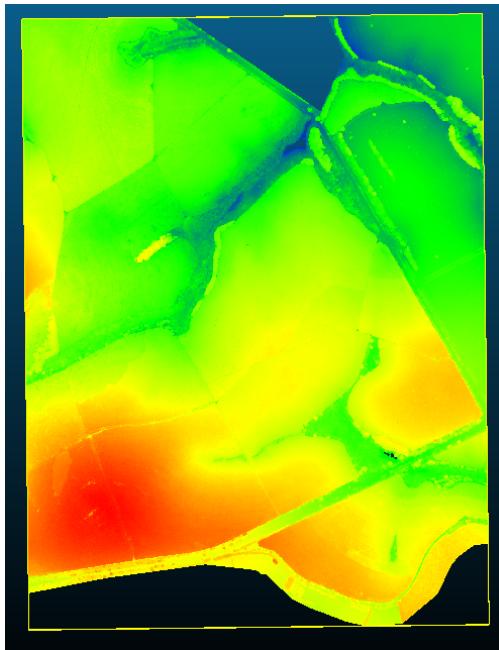


Figure 8: Nuvens de pontos LiDAR pré-processadas

- ii. As nuvens foram baixadas pelo seguinte link: https://github.com/FlorestaR/dados/blob/main/5_LIDARF/Modelo/CLOUDS/

- . Download e leitura dos dados em Shapefile
- i. Os shapes foram baixados pelo seguinte link: https://github.com/FlorestaR/dados/blob/main/5_LIDARF/Modelo/SHAPEs

Fazenda	A
ID_Talhao	301
CODTALHAO	52I214
CHAVE	52I214_6
CHAVE2	P301410062
NUMPARCELA	6
NUMMEDICAO	2
ANOREF	2013
DATAREALIZ	4/18/2013
IDINV	3,70000
AREAPARCEL	399,73000
LAT_UTM23S	206632,09800
LONG_UTM23	7357127,11200
MHDOM	19,37000
VTCC	199,85000
VCCC	196,85000
SUBTALHAO	301d

Figure 9: Dados contidos nas parcelas inventariadas

. Clipagem da nuvem para eliminação de dados indesejados (mostrar nuvem antes e depois)

Antes da clipagem: 68mi pontos

Após clipagem: 18mi pontos

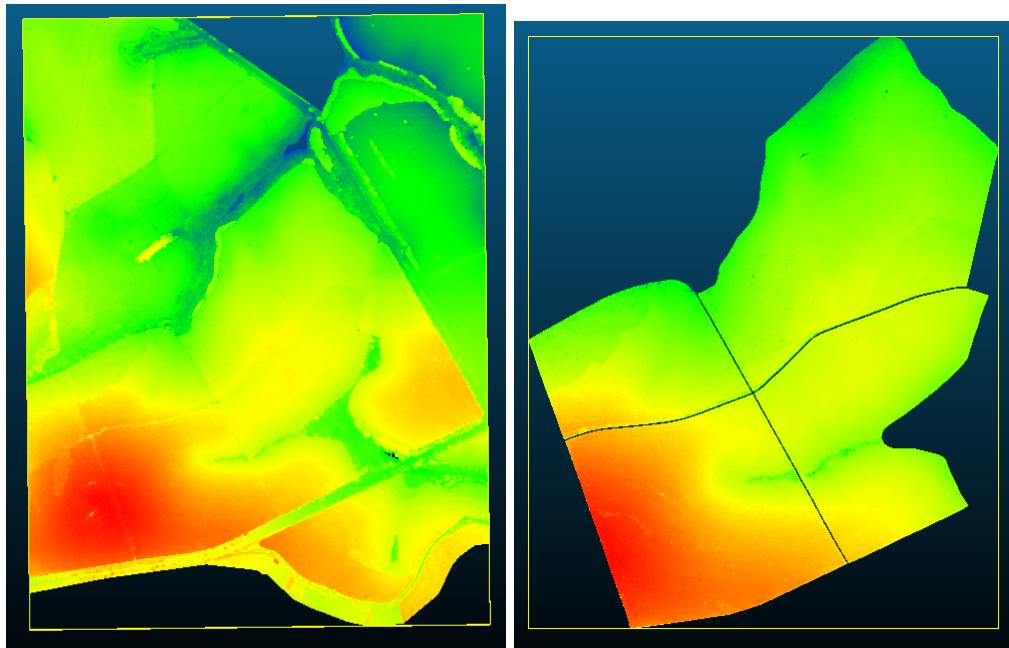


Figure 10: Comparativo entre as nuvens de pontos antes e após a clipagem

. Classificação

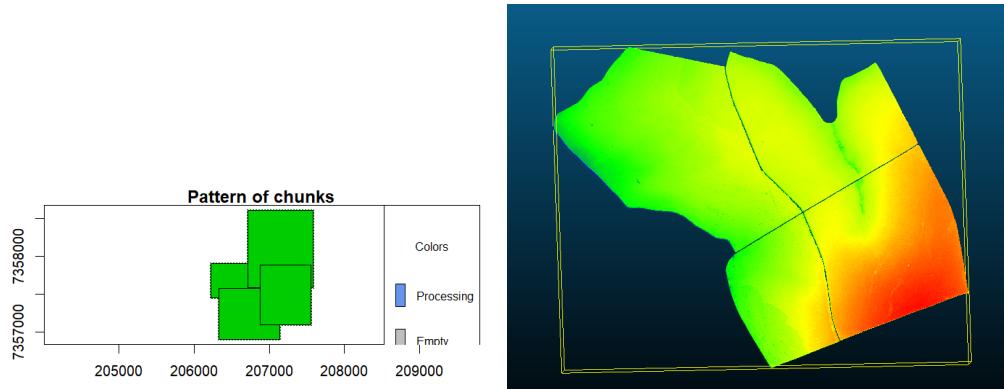


Figure 11: Processamento e resultado da classificação de solo das nuvens LiDAR

. Normalização i. A etapa de normalização tem por finalidade nivelar toda a nuvem e é um passo que está diretamente correlacionado à classificação do solo.

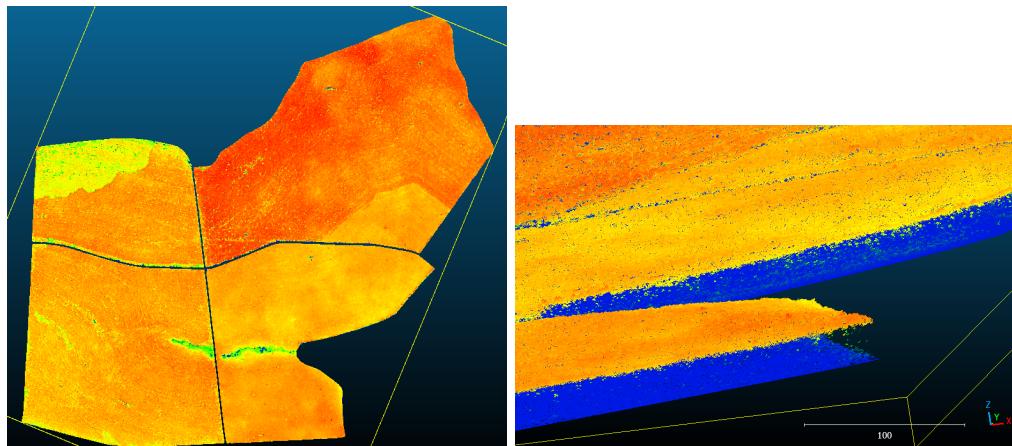


Figure 12: Resultado da normalização das nuvens

10

. Recorte da nuvem em tiles 300x300m O recorte da nuvem em tiles menores tme a função de facilitar o processamento dos dados, tornando-o mais rápido e dinâmico

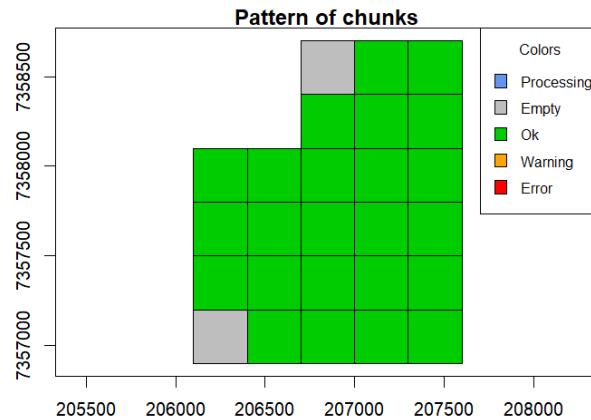


Figure 13: Retile da nuvem em quadrados de 300x300m

11. Cálculo de métricas para cada voxel (explicar voxel)

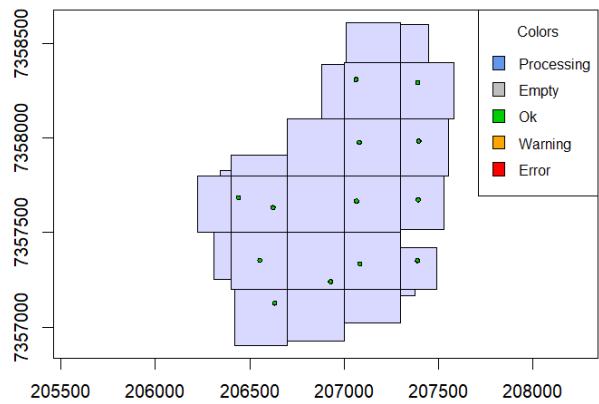


Figure 14: Cálculo das métricas para cada voxel

. Seleção de um grupo de métricas para estudo de correlação

	VTCC	MHDOM	IDINV	Zmean	zq45	zq75	zq95	zpcum2	zpcum4	zpcum6	pzabovezmean	pzabove2
	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>	<db>
1	247.	22.2	5.2	12.6	13.1	23.1	24.2	1.08	1.08	1.55	55.1	55.2
2	204.	19.9	5.2	11.8	2.98	21.9	22.7	8.81	8.94	9.65	54.0	56.3
3	285.	27.4	5.2	13.2	19.8	22.1	23.7	1.14	1.34	1.59	61.0	61.3
4	249.	24.5	5.2	14.1	20.4	22.7	24.2	1.21	3.21	3.94	63.8	64.7
5	335.	32.7	5.2	12.5	0	26.0	27.2	7.19	8.62	16.8	49.3	51.5
6	307.	28.1	5.2	12.3	0.37	24.8	25.9	9.33	9.50	10.3	49.9	51.0
7	200.	19.4	3.7	13.2	19.8	22.1	23.7	1.14	1.34	1.59	61.0	61.3
8	219.	19.9	3.7	14.1	20.4	22.7	24.2	1.21	3.21	3.94	67.8	64.7
9	155.	14.8	3.7	10.7	15.3	18.8	20.8	2.47	3.20	3.52	57.9	59.3
10	254.	23.4	3.7	14.7	21.4	23.6	25.5	4.86	5.56	5.56	63.1	65.3
11	247.	22.1	3.7	13.7	20.3	23.0	24.5	10.6	11.0	11.2	60.1	63.8
12	238.	23.9	5.2	14.0	21.2	22.9	24.2	2.04	2.19	2.55	62.0	62.8
13	276.	25.8	5.2	15.9	23.2	24.4	25.2	2.55	2.55	2.55	65.9	66.4

Figure 15: Tabela com as métricas escolhidas

. Análise da correlação por meio de gráfico (falar do gráfico)

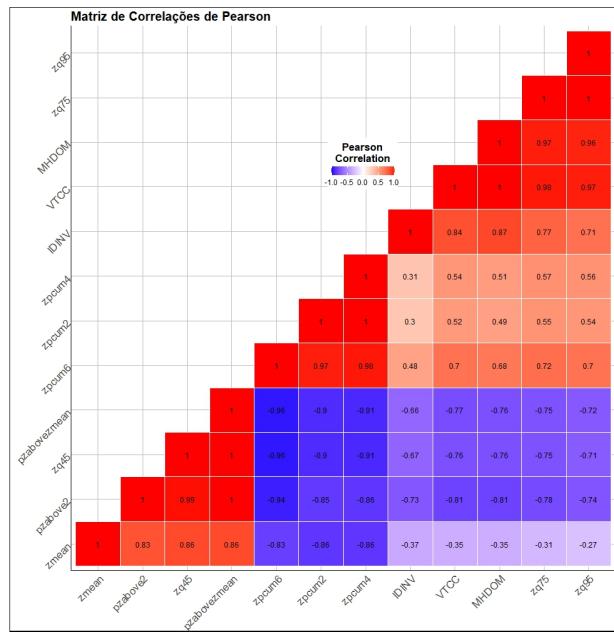


Figure 16: Resultado da análise de correlação

. Determinação do modelo de predição

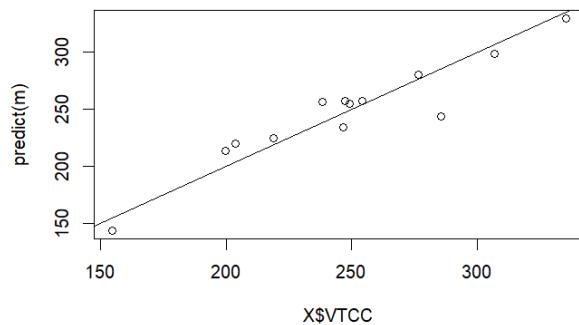


Figure 17: nao sei

15

. Criação de mapas raster e gráficos

6 Fluxograma e etapas Tripla amostragem

Composta por 3 fases, s0, s1 e s2:

- O princípio básico é o de que as variáveis explanatórias derivadas das informações auxiliares estão disponíveis em duas frequências diferentes. A fase s0 fornece informações sobre toda a área, enquanto s1 possui dados adicionais de amostras de s0. Logo, a partir da informação terrestre coletada de um número x de parcelas de campo (1 camada de informação) é possível aferir sobre informações adicionais para outras parcelas a partir do uso de preditores (2 camadas de informação) e, por fim, o LiDAR coleta dados sobre todas as parcelas (3 camadas de informação).

- Logo, a motivação por trás da TA é a de que a gama de informações de s1 adiciona alto poder preditivo às variáveis disponíveis para todas as parcelas da área (s0)

Journal of Statistical Software

5

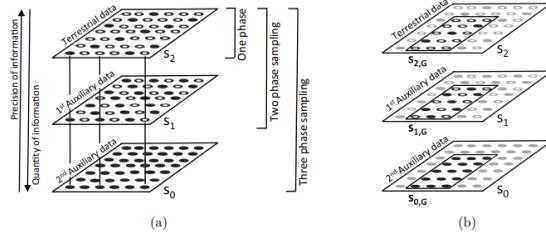


Figure 2: (a) Concept of multiphase sampling. The square represents the forest area for which an inventory is being conducted. The points denote the sample locations x . Filled points indicate available information. (b) Illustration of the small area estimation problem.

Figure 18: Esquema TA com Lidar

Estimativas de pequenas áreas (small area estimation)

- Para sub áreas da floresta onde há pouca informação terrestre, como em G na figura acima. Para essas áreas, o uso da amostragem multifásica pode ser mais eficiente, já que utilizam um número reduzido de parcelas em campo para chegar à mesma precisão da ACE e ACS. Por outro lado, a sub área em questão pode ser pequena demais para justificar a adoção de um modelo de regressão separado, uma vez que este pode resultar em um intervalo de confiança indesejadamente abrangente;

- A ideia, então, é a de utilizar toda a riqueza de informações presente nas amostras de S2 para ajustar o modelo (equação utilizada para o cálculo de uma variável de interesse) e aplicá-lo para a sub área em questão.

- O potencial viés que surge da aplicação do modelo na sub área é corrigido pelo uso de modelos residuais empíricos derivados da área amostrada em campo. (não entendi direito essa parte) - trecho no texto: The potential bias of applying that model in G is then corrected for by using the empirical model residuals derived from that small area.

- Caso não existam parcelas de campo na sub área, então deve-se aceitar o viés na estimativa e no modelo. São essas as estimativas sintéticas, mas apesar do viés é possível calcular a sua variação.

DESIGN-BASED VS MODEL-DEPENDENT APPROACH (ver dps)

- Model-dependent approach: as parcelas amostrais são fixas e as observações retiradas desses locais são assumidas como variáveis aleatórias, assim como a floresta assume o papel de ser o meio realizador desse processo estocástico (????? o que isso significa?). Embora os locais das parcelas possam ser arbitrariamente escolhidos, o modelo deve escrever adequadamente o processo estocástico, a fim de garantir resultados parciais. (????)

- Os estimadores do forestinventory se baseiam na design-based approach. Design-based approach baseia-se na randomização dos locais de amostra de forma uniforme e independente. Logo, a floresta por si só e qualquer valor de densidade local de x (pertencente a) F (floresta) são fixos e não um resultado de um processo estocástico (estudar). (????)

- Na estrutura do Model-dependent approach as parcelas são fixas e resultam em um processo estocástico GPT: what is the difference between design-based approach and model-dependent approach?