
Estatística: Aplicação ao Sensoriamento Remoto

SER 204

Introdução

Camilo Daleles Rennó

camilo.renno@inpe.br

acesso do conteúdo do curso em [Bibdigital do INPE](#) ou [GitHub](#)

Estudando um problema...

Em geral, numa pesquisa estamos interessados em estudar alguns aspectos relacionados a algum fenômeno real ou hipotético (modelo) tentando compreender os diversos resultados observados ou possíveis de serem obtidos



População, Universo ou Espaço Amostral

Quase sempre é impossível analisar todos os dados relacionados a esses aspectos, o que em estatística chamamos de **população**, universo ou espaço amostral

De fato, a população representa o conjunto de todos os resultados possíveis de um experimento e é designado por U , S ou Ω

Por exemplo, num estudo sobre qual é o papel do fogo no processo de desmatamento na Amazônia, quem faz parte da população estudada? Ou seja, qual é o escopo do meu estudo ou onde ele se aplica?

Inclui todos os desmatamentos já ocorridos? De qualquer tamanho? Em qualquer lugar? Considera-se fogo natural, “acidental” e criminoso?

Assim, a definição do escopo do estudo é fundamental para entendermos sua abrangência e suas limitações

Probabilidade

Diversos fenômenos que estudamos não são perfeitamente previsíveis, ou seja, o resultado observado não poderia ser conhecido antecipadamente

Esses resultados podem representar a combinação de inúmeros processos, alguns deles desconhecidos, ou então esses processos podem ter componentes aleatórias

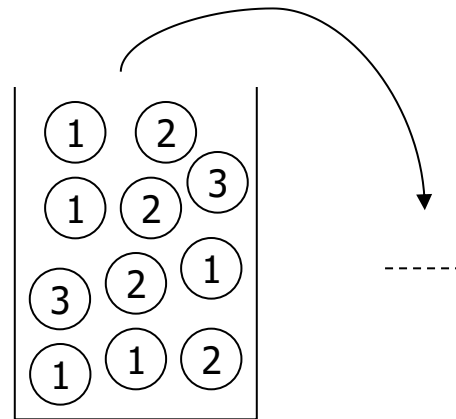
No entanto, é possível verificar que certos resultados são mais facilmente obtidos do que outros.

Por exemplo, se 70% de um mapa é representado pela classe Floresta e sortearmos um ponto qualquer neste mapa, podemos afirmar que este ponto será de Floresta ou apenas temos a indicação de que é mais fácil (mais provável ou com maior **probabilidade**) que este ponto seja da classe Floresta?

Probabilidade

Quando conhecemos ou temos acesso a todos os elementos da população, podemos caracterizar perfeitamente as propriedades probabilísticas desta população

Exemplo: uma bola é escolhida ao acaso de uma urna



$$P(1) = 5/11$$

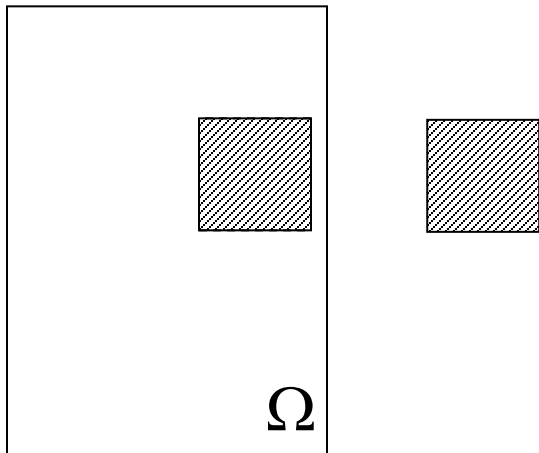
$$P(2) = 4/11$$

$$P(3) = 2/11$$

$$P(evento) = \frac{\#eventos\ favoráveis}{\#eventos\ possíveis}$$

Amostragem

Em geral, toda a análise é feita com base numa pequena parte da população que chamamos de **amostra**



Questões básicas:

- a) Quantas observações devo ter para garantir que minha análise produza resultados confiáveis?
 - . tamanho de amostra ideal
- b) Como devo proceder a coleta destas informações?
 - . técnicas de amostragem
- c) Qual é o procedimento estatístico mais adequado para analisar meus dados?
 - . tipo de dados e métricas representativas
 - . hipóteses a serem testadas
 - . relacionamento entre observações

Planejamento

Idealmente a escolha da análise estatística deveria ser feita no início de seu projeto, antes mesmo de se coletar quaisquer dados

As análises estatísticas têm diferentes pré-requisitos relacionados ao tipo de dado analisado, tamanho da amostra e natureza de relacionamento desses dados. Conhecer a abordagem estatística previamente permitirá que você planeje a forma como você deve coletar seus dados

Cuidado! Se você sair coletando dados para daí procurar uma maneira de realizar a análise dos mesmos, você pode ao final concluir que não dispõe exatamente do que você precisa para responder a sua questão científica.



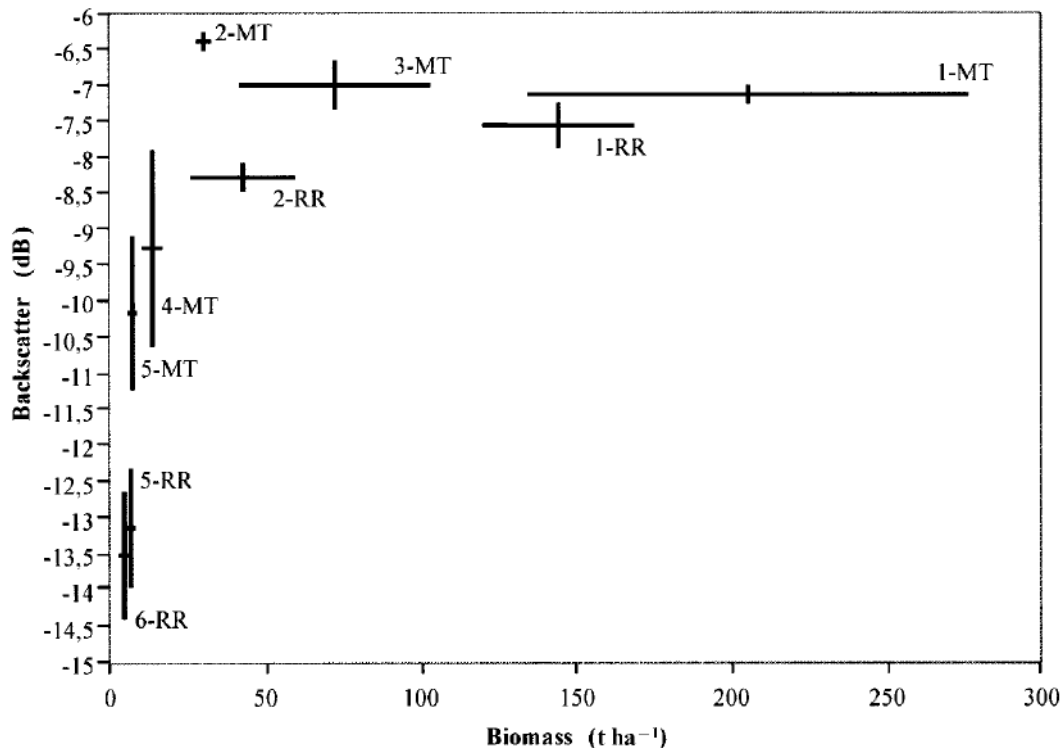
Definição Clássica de Estatística

Estatística é a ciência que investiga os processos de obtenção, organização e análise de dados sobre uma população, e os métodos de tirar conclusões ou fazer previsões com base nesses dados.

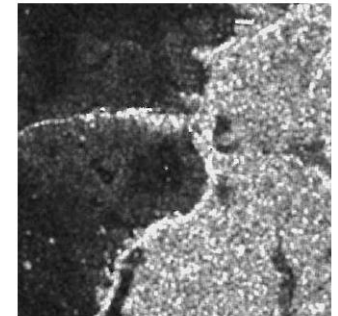
Estatística Descritiva X **Estatística Inferencial**

Para que Estatística em SR?

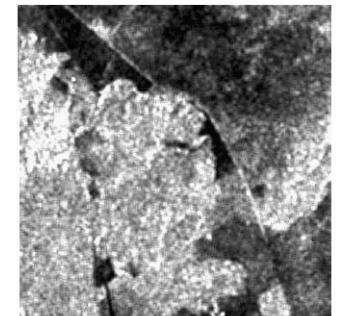
- caracterizar um fenômeno
ex: qual o retroespalhamento de áreas florestais em dados JERS?



Mucajai/Roraima



Comodoro/Mato Grosso

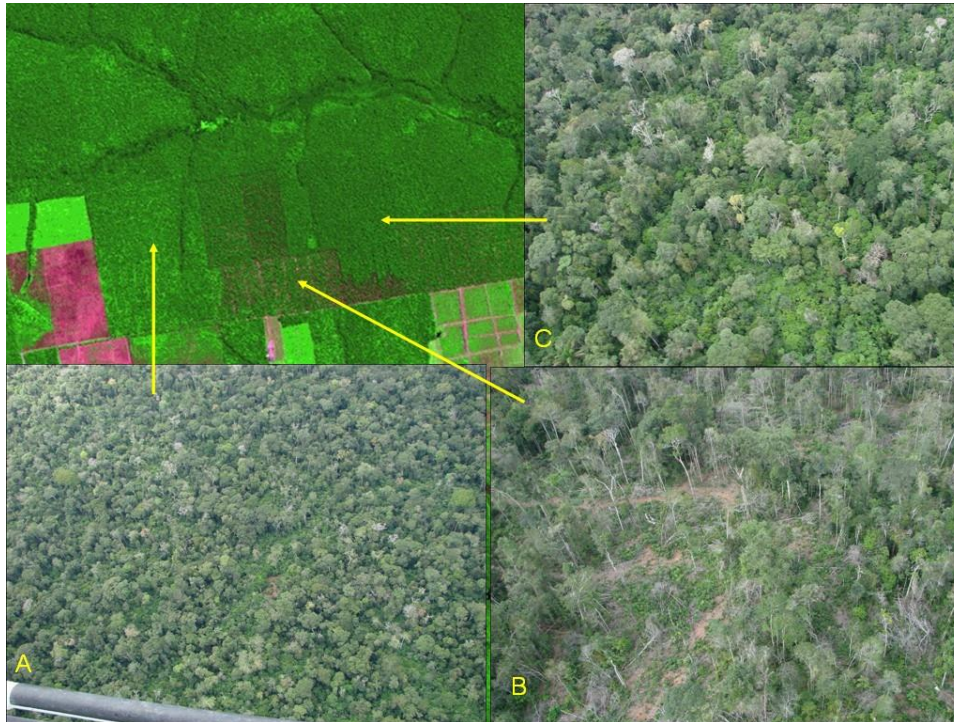


Fonte: Santos et al. Savanna and tropical rainforest biomass estimation and spatialization using JERS-1 data. Int. J. Remote Sensing, 23(7):1217-1229. 2002

Para que Estatística em SR?

- comparar fenômenos

ex: é possível distinguir diferentes níveis de degradação em florestas tropicais a partir de imagens Landsat/TM?

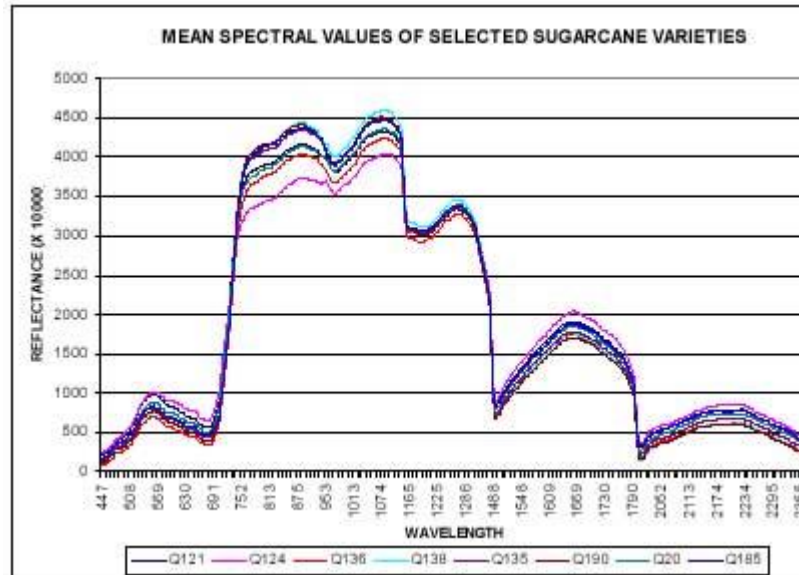


Fonte: <http://www.obt.inpe.br/degrad/>

- . quais as melhores métricas e bandas espectrais a serem usadas?
- . quantos níveis de degradação podem ser detectados?
- . qual a melhor técnica ou procedimento para se fazer essa detecção?

Para que Estatística em SR?

- encontrar padrões
ex: qual o comportamento espectral típico da cana-de-açúcar?



Fonte: <http://www.a-a-r-s.org/aars/proceeding/ACRS2004/Papers/HSS04-1.htm>

Para que Estatística em SR?

- selecionar fatores mais importantes para explicar um fenômeno
ex: quais bandas espectrais e índices de vegetação que melhor explicam a variação de Biomassa e Carbono orgânico em macrófitas?

Tabela 9 - Matriz de correlação entre as bandas espectrais e variáveis biofísicas em cada sensor. Em negrito, as mais altas correlações.

		Biom	Carb	Azul	Verde	Verm	NIR	RE5	RE6	RE7	RE8A	SWIR1	SWIR2
Planet Scope	Biom	1											
	Carb	0.99	1										
	Azul	-0.50	-0.51	1									
	Verde	-0.27	-0.28	0.90	1								
	Verm	-0.36	-0.35	0.93	0.97	1							
	NIR	0.60	0.61	0.76	-0.49	-0.55	1						
Sentinel 2-A	Azul	0.07	0.07	1									
	Verde	0.33	0.35	0.84	1								
	Verm	0.04	0.06	0.97	0.79	1							
	NIR	0.45	0.47	0.59	-0.12	-0.67	1						
	RE5	0.21	0.02	0.87	0.93	0.83	0.22	1					
	RE6	0.48	0.50	0.68	-0.24	-0.75	0.96	0.33	1				
	RE7	0.43	0.46	0.68	-0.24	-0.75	0.96	0.33	0.98	1			
	RE8A	0.46	0.48	0.68	-0.24	-0.74	0.95	0.35	0.98	0.98	1		
	SWIR1	0.42	0.41	0.72	0.80	0.61	0.01	0.76	0.16	0.16	-0.18	1	
	SWIR2	0.22	0.19	0.86	0.78	0.78	0.33	0.78	0.47	0.47	-0.50	0.93	1

*Biom: Biomassa seca; Carb: carbono orgânico; Verm: banda correspondente ao vermelho do espectro EM; NIR: banda correspondente ao Infravermelho próximo; RE: banda correspondente à faixa de borda vermelha; SWIR: banda correspondente ao Infravermelho de ondas curtas.

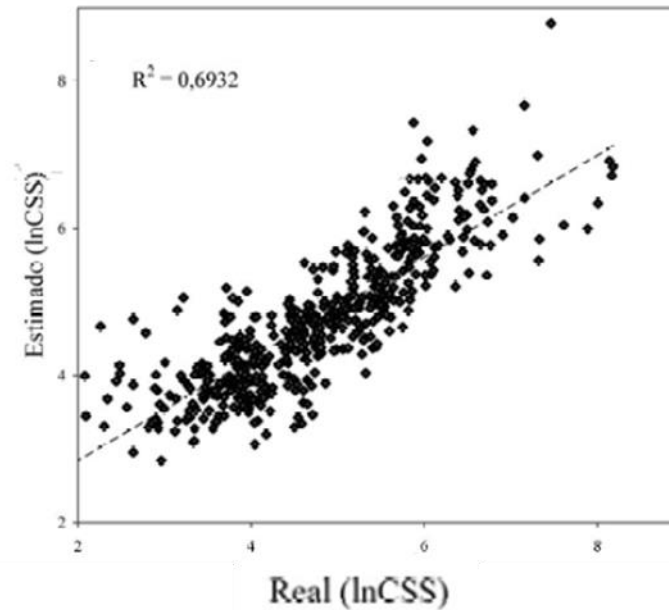
Tabela 10. Matriz de correlação entre os índices de vegetação e variáveis biofísicas em cada sensor. Em negrito, as mais altas correlações.

		Biom	Carb	NDVI	NDAVI	WAVI	SPRI	CO ₂ Flux	CO ₂ flux NDAVI
Planet Scope	Biom	1							
	Carb	0.99	1						
	NDVI	0.55	0.55	1					
	NDAVI	0.57	0.58	0.99	1				
	WAVI	0.57	0.58	0.99	1.00	1			
	SPRI	0.66	0.68	0.69	0.76	0.76	1		
	CO ₂ Flux	0.59	0.59	0.99	1.00	1.00	0.77	1	
	CO ₂ Flux NDAVI	0.61	0.62	0.97	0.99	0.99	0.83	0.99	1
Sentinel- 2A	NDVI	0.24	0.26	1					
	NDAVI	0.25	0.28	0.99	1				
	WAVI	0.25	0.28	0.99	1	1			
	SPRI	0.19	0.22	0.94	0.97	0.97	1		
	CO ₂ Flux	0.66	0.65	0.16	0.10	0.10	-0.04	1	
	CO ₂ Flux NDAVI	0.30	0.33	0.83	0.86	0.86	0.83	0.15	1

Para que Estatística em SR?

- encontrar relações matemáticas entre fenômenos
ex: qual modelo matemático que melhor estima a concentração de sólidos em suspensão de corpos d'água na Amazônia a partir da reflectância de superfície em dados do Landsat/TM?

Variável	Coefficiente
Intercepto	-1,80411
$\frac{B4 + B3}{B2 + B1}$	3,41598
B4/B3	2,81481
B5	-0,32422
B3/B2	2,2984
B2/B1	1,8412
B3/B1	-1,36105



Fonte: <http://mtc-m19.sid.inpe.br/col/sid.inpe.br/mtc-m19/2013/02.26.19.52/doc/publicacao.pdf>

Para que Estatística em SR?

- comparar ou avaliar classificações/mapeamentos/modelos
ex: qual método de classificação resulta no melhor mapeamento de uma área urbana?



Classificação MAXVER



Imagem IKONOS fusionada



Classificação orientada a objetos

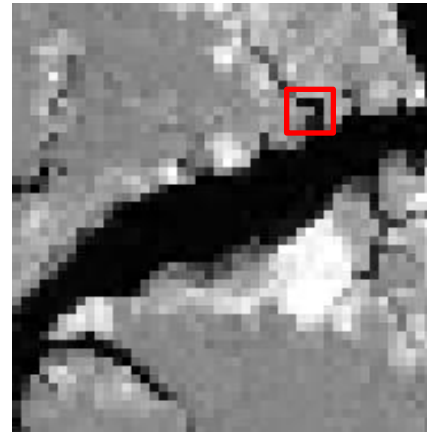
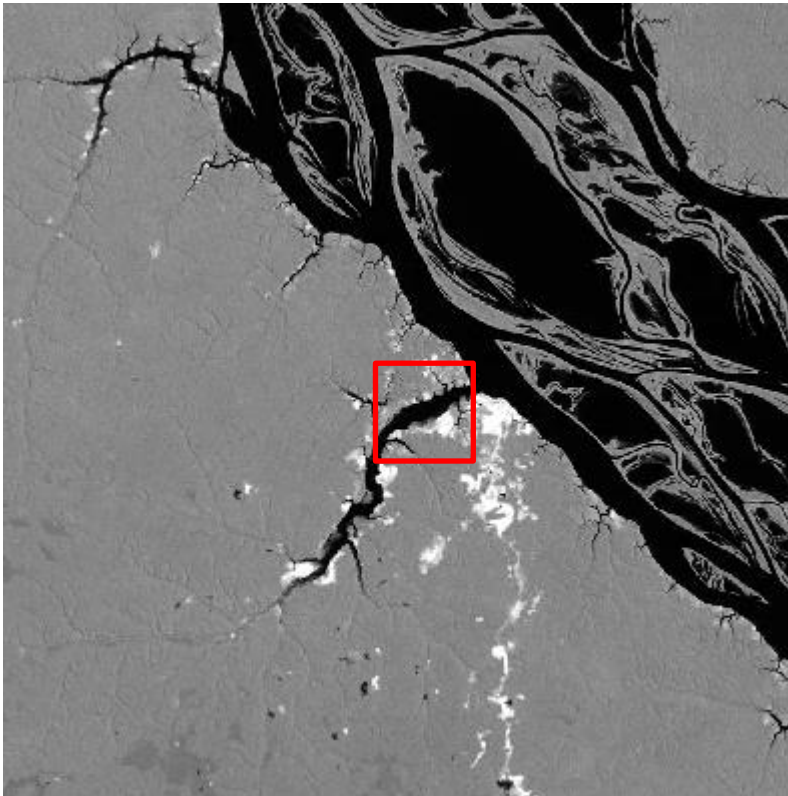
Legenda

■	Vegetação
■	Solo exposto
■	Telha cerâmica
■	Telha alumínio
■	Concreto/amianto claro
■	Concreto/amianto escuro
■	Pavimentação
■	Água
■	Sombra

Fonte: <http://martel.sid.inpe.br/col/ltid.inpe.br/sbsr/2004/11.19.17.58/doc/4217.pdf>

Estatística em SR

O que difere a estatística aplicada no Sensoriamento Remoto daquela aplicada em outras áreas do conhecimento?

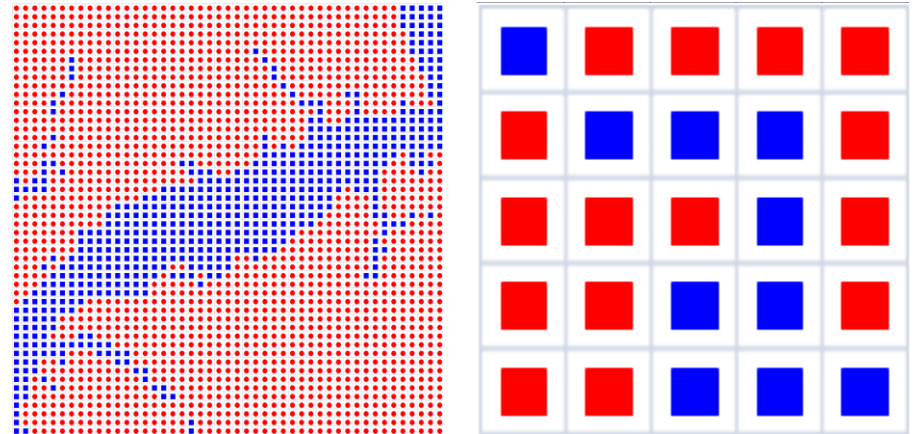
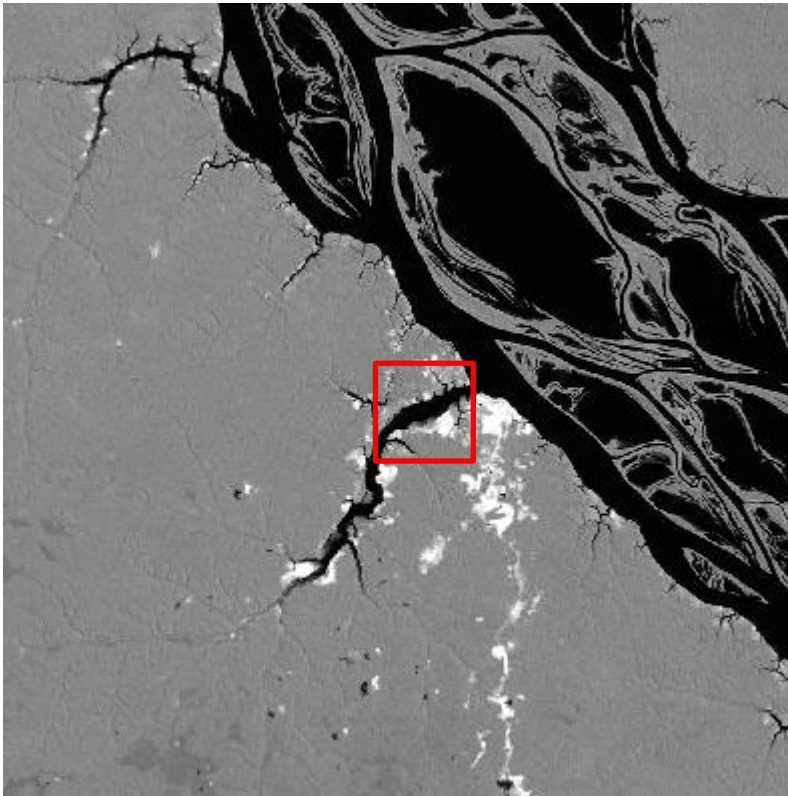


99	174	174	187	162
163	16	7	16	156
175	148	142	9	152
149	137	32	23	155
139	102	28	18	99

Uma imagem é uma grade ou matriz, onde cada número representa o valor de uma grandeza (radiância, reflectância, retroespalhamento, etc) de um elemento de resolução (*pixel* ou célula). A visualização é feita associando-se a cada número um nível de cinza (ou uma cor).

Estatística em SR

O que difere a estatística aplicada no Sensoriamento Remoto daquela aplicada em outras áreas do conhecimento?

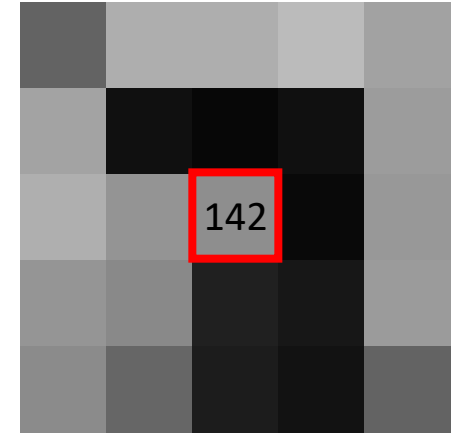
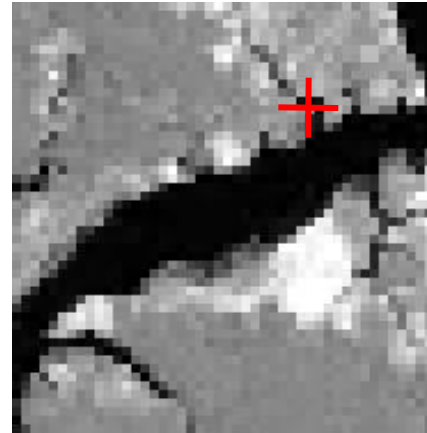
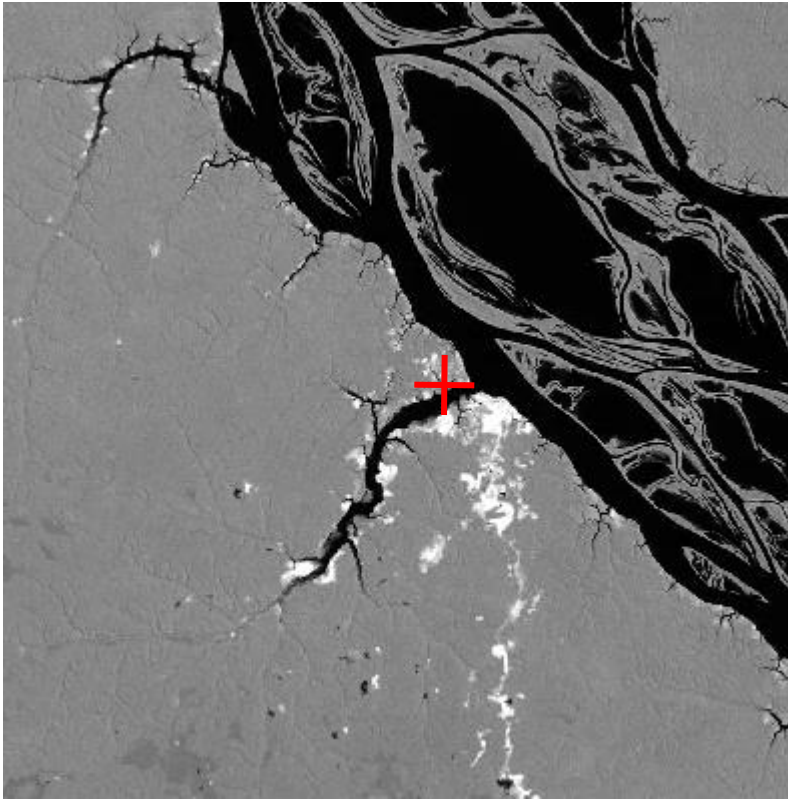


Numa imagem classificada, cada elemento de resolução está associado a uma classe (ou rótulo) de acordo com um conjunto de regras.

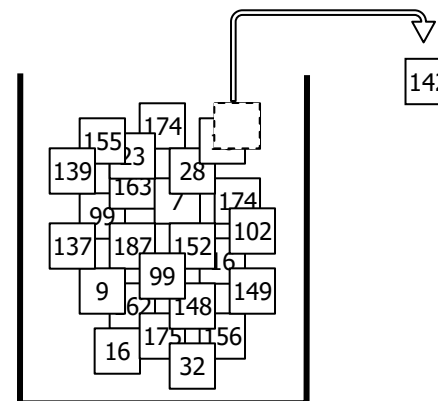
fatiamento

■ ≤ 100 ■ > 100

Estatística em SR (imagens)



Desconsiderando-se as relações espaciais, este sorteio seria equivalente a um sorteio numa urna!



Sorteio aleatório de 1 ponto...