

Minicurso 1 (MC1): Modelagem Estatística de Imagens Digitais

Bruna Palm (Blekinge Institute of Technology)
Fabio Bayer (UFSM)

Outubro, 2025

Sumário

- 1 O que é uma Imagem?
- 2 Imagem Digital
- 3 Processamento de Imagens
- 4 Operações Morfológicas
- 5 Sensoriamento Remoto
- 6 Modelos Propostos
 - Detector de Mudanças — Distribuição Rayleigh
 - Modelo de Regressão Rayleigh
 - Correção de Viés no Modelo de Regressão Rayleigh
 - Estimadores Robustos no Modelo de Regressão Rayleigh
 - Distribuição Rayleigh Inflacionada
 - Modelo de Regressão Rayleigh Inflacionado
 - Model ARMA Rayleigh Inflacionado
 - Previsão
 - Model ARMA 2-D Rayleigh
 - Modelos Regressão para a Distribuição Kumaraswamy Modificada Inflacionada
 - Modelo ARMA 2-D Kumaraswamy

Sobre mim

- Brasileira, doutora em Estatística pela UFPE (2020).
- Bacharela em Estatística pela UFSM (2014).
- *Associate Professor* na Blekinge Institute of Technology (BTH), Suécia.
- Atuação principal em modelagem estatística aplicada à análise de imagens e séries temporais.
- Experiência em projetos nas áreas de processamento de sinais, médica, previsão hidrológica e otimização ferroviária.

Sobre a BTH

- A *Blekinge Institute of Technology* (BTH) é uma universidade pública sueca localizada em Karlskrona.
- Reconhecida pelo foco em sustentabilidade, inovação e integração entre tecnologia e gestão.
- Oferece programas nas áreas de tecnologia, TI e ciências aplicadas, com forte colaboração com a indústria.



Sobre Fábio Bayer

- Doutor em Estatística pela UFPE (2011).
- Professor Associado do Departamento de Estatística da UFSM.
- Bolsista de Produtividade em Pesquisa do CNPq desde 2015.
- Professor nos Programas de Pós-Graduação em Matemática (UFSM), Engenharia de Produção (UFSM) e Estatística (UFRGS).
- Atuação em modelos de regressão, modelos dinâmicos e processamento de sinais.

Sobre a UFSM

- Primeira universidade federal fundada fora de uma capital.
- Rotineiramente posicionada entre as posições 11 e 17 (nacionalmente) em rankings de instituições científicas.
- Departamento de Estatística renovado e motivado!
- Siga @statufsm no Instagram.



Sobre notação

Tabela: Referência cruzada de termos estatísticos aos termos de engenheiros

Termos na Estatística	Termos na Engenharia
Amostra	Sinal
Observação	Amostra
Indexação: y_i , $i = 1, 2, \dots n$	$y[n]$, $n = 1, 2, \dots N$
Teste de hipóteses	Detector
Hipótese nula (H_0)	Hipótese de apenas ruído
Hipótese alternativa (H_1)	Hipótese de sinal + ruído
Erro Tipo I (decidir H_1 quando H_0 verd.)	Falso alarme (FA)
Erro Tipo II (decidir H_0 quando H_1 verd.)	Falha
Probabilidade de erro Tipo I (α)	Probabilidade de falso alarme (P_{FA})
Probabilidade de erro Tipo II (β)	Probabilidade de falha (P_M)
Poder do teste ($1 - \beta$)	Probabilidade de detecção (P_D)
Região crítica	Região de decisão de sinal presente

O que é uma Imagem?

- No dia a dia, chamamos de imagem qualquer representação visual: foto, mapa, radiografia.
- Matematicamente, uma imagem é uma função que associa a cada posição (x, y) um valor de intensidade ou cor:

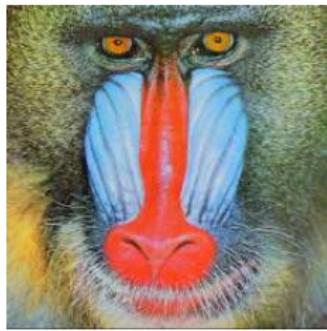
$$f(x, y) \rightarrow \text{intensidade ou cor do ponto na posição } (x, y).$$

- É como uma grande planilha de números: cada célula corresponde a um ponto do espaço.

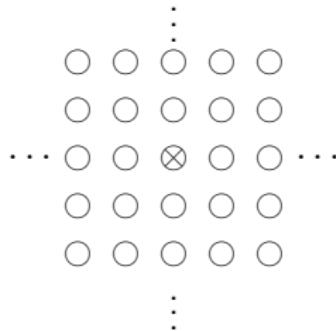
De Função Contínua a Imagem Digital

- Na natureza, $f(x, y)$ é contínua, mas no computador precisamos discretizar.
- Amostragem: dividimos o espaço em pixels, como uma grade.
- Cada pixel armazena um valor de intensidade ou cor.
- Quantização: arredondamos os valores para um conjunto finito (0–255 em tons de cinza, por exemplo).
- O resultado é uma matriz de números — a imagem digital.
- Ou seja, uma imagem digital é uma representação discreta de uma cena.

O que é uma Imagem Digital?



(a) Imagem Digital



(b) Pixels

Formação da Imagem Digital

- A formação começa com um objeto no mundo real.
- Uma onda eletromagnética incide na cena (passivo ou ativo).
- A energia refletida é captada por um sensor (câmera, satélite).
- Forma-se uma imagem analógica no plano da câmera.
- Essa imagem é digitalizada:
 - ▶ **Amostragem:** divisão em uma grade de pixels.
 - ▶ **Quantização:** atribuição de um valor a cada pixel.

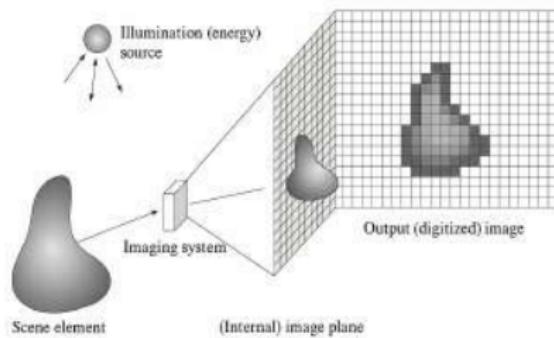
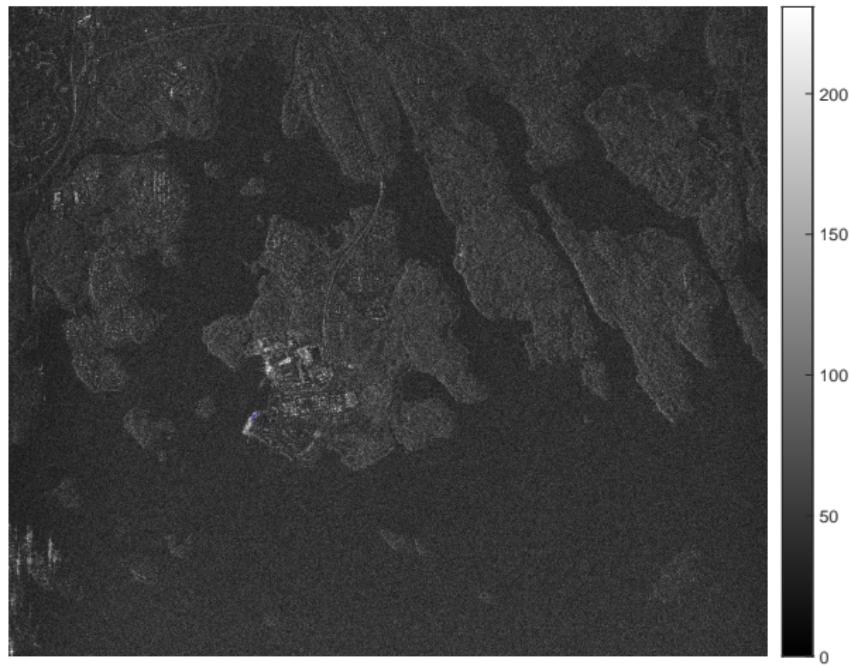


Image from: Digital Image Processing, Rafael C. Gonzalez and Richard E. Woods, 2018

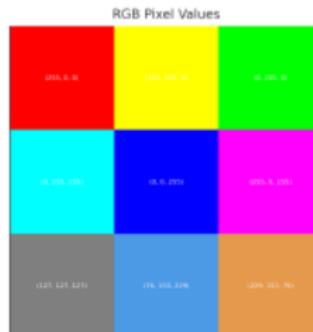
Um exemplo de Imagem Digital (radar)?



Representação de Pixels

- **Tons de cinza:** um valor por pixel ($0 =$ preto, $255 =$ branco), que pode representar diferentes características.
- **RGB:** três valores por pixel (vermelho, verde, azul).

Valores de Pixel em Tons de Cinza (0-255)



Sistema de Coordenadas e Profundidade de Bits

- A imagem é uma matriz 2D, cada pixel guarda “um” valor.
- Profundidade de bits: número de valores possíveis por pixel.
- 8 bits (grayscale): $2^8 = 256$ tons de cinza.
- RGB 8 bits: $256^3 = 16,777,216$ combinações de cores.
- Profundidades maiores (16 bits) permitem mais gradações.

Formação de Cores em Imagens RGB

- Cada canal contribui para a cor final por mistura aditiva:

$$\text{Cor} = R \cdot \text{Vermelho} + G \cdot \text{Verde} + B \cdot \text{Azul}$$

- Exemplos:

- ▶ $(255, 0, 0) \rightarrow \text{Vermelho}$
- ▶ $(0, 255, 0) \rightarrow \text{Verde}$
- ▶ $(0, 0, 255) \rightarrow \text{Azul}$
- ▶ $(255, 255, 0) \rightarrow \text{Amarelo}$
- ▶ $(255, 255, 255) \rightarrow \text{Branco}$
- ▶ $(0, 0, 0) \rightarrow \text{Preto}$

Por que Processar Imagens?

- Aquisição: melhorar contraste, balanço de cores, reconstrução 3D.
- Realce: remoção de ruído, aumento de nitidez.
- Armazenamento e transmissão: compressão eficiente.
- Extração de informação: detecção de faces, leitura de códigos de barras.

Exemplos de Aplicações

- Detecção e desfoque de rostos para privacidade.
- Segmentação de órgãos em imagens médicas (ex.: tomografia).
- Classificação de uso do solo em sensoriamento remoto.
- Realidade aumentada e visão computacional.

Por que uma Abordagem Estatística?

- Assim como em séries temporais ou regressão, os pixels contêm variabilidade e ruído.
- Uma imagem real raramente é perfeita: há ruído, limitações do sensor, iluminação variável.
- A estatística permite:
 - ▶ Modelar o comportamento esperado de pixels ou regiões.
 - ▶ Separar padrão de ruído.
 - ▶ Fazer inferências: detectar mudanças, classificar áreas, segmentar objetos.

Operações Morfológicas

- São operações aplicadas a imagens binárias ou em tons de cinza, baseadas na forma dos objetos.
- Utilizam um **elemento estruturante** (kernel), que é uma pequena matriz que “varre” a imagem (convolução).
- Principais operações:
 - ▶ **Erosão**: “encolhe” os objetos.
 - ▶ **Dilatação**: “expande” os objetos.
 - ▶ **Abertura**: erosão seguida de dilatação — remove pixels isolados (ruídos).
 - ▶ **Fechamento**: dilatação seguida de erosão — preenche pequenos “buracos”.

Definições Matemáticas

Erosão: Reduz o objeto. Mantém o pixel se **todo** o elemento estruturante B está contido no objeto A .

$$A \ominus B = \{z \mid B_z \subseteq A\}$$

Dilatação: Expande o objeto. Mantém o pixel se **alguma** parte de B intersecta A .

$$A \oplus B = \{z \mid (B)_z \cap A \neq \emptyset\}$$

Abertura: Remove pequenos objetos e suaviza contornos.

$$A \circ B = (A \ominus B) \oplus B$$

Fechamento: Preenche buracos e conecta pequenos espaços.

$$A \bullet B = (A \oplus B) \ominus B$$

Exemplos Visuais



Figura: Exemplo ilustrativo de operações morfológicas

- (a) Imagem original
- (b) Erosão: objetos (pixels brancos) ficam menores
- (c) Dilatação: objetos crescem
- (d) Abertura: remove pequenos pontos isolados
- (e) Fechamento: preenche pequenos buracos

Intuição Prática

- **Erosão:** útil para remover pequenos ruídos ou separar objetos conectados.
- **Dilatação:** útil para destacar objetos ou conectar regiões quebradas.
- **Abertura:** “limpa” a imagem removendo objetos muito pequenos.
- **Fechamento:** preenche falhas e buracos em objetos.

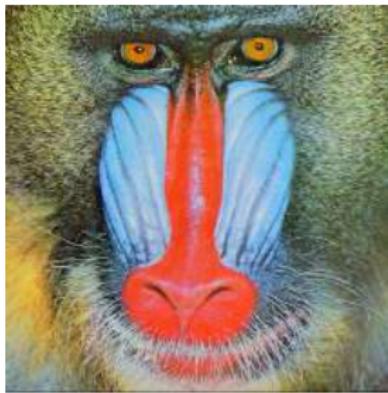
Dica Intuitiva

Pense no elemento estruturante como um carimbo que “testa” cada pixel da imagem.

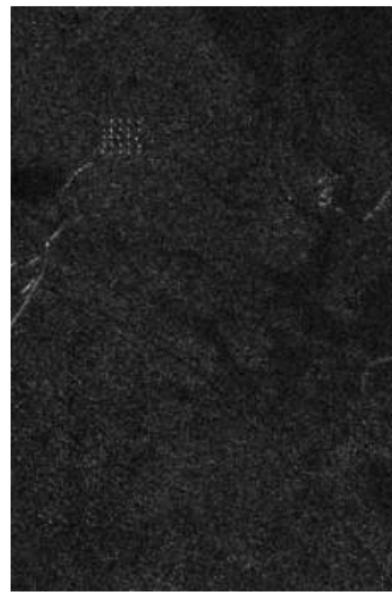
Sensoriamento Remoto - Tipos de Sensores

- Passivos (Ópticos): captam radiação refletida (ex.: foto, Landsat, Sentinel-2). Com eles, enxergamos o que os olhos veem.
- Ativos (Radar): emitem o sinal e captam o eco (ex.: Sentinel-1, RADARSAT). Com eles, detectamos coisas além de certas barreiras físicas (por exemplo, atravessam nuvens e copas de árvores). Ferramenta poderosa para monitoramento da Terra.

Sensoriamento Remoto - Tipos de Sensores



Passivo



Ativo

Dados SAR - Ativo

- SAR = Synthetic Aperture Radar.
- Radar ativo: emite micro-ondas e registra o eco.
- Produz imagens de amplitude (cada pixel é um numero complexo).
- Característica principal: presença de speckle (ruído multiplicativo).

Índice NDVI - Passivo

- NDVI = Normalized Difference Vegetation Index.
- Calculado a partir de bandas Vermelho (Red) e Infravermelho Próximo (NIR):

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}.$$

- Mede “vigor” da vegetação (valores próximos de 1 indicam vegetação saudável).

Distribuição de Rayleigh em SAR

- Modela a intensidade (amplitude) do sinal em regiões homogêneas.
- Função densidade:

$$f(z; \sigma) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad z > 0.$$

- Parâmetro σ : relacionado ao brilho médio da região.

Por que a Distribuição de Rayleigh em SAR?

- Justificativa teórica:

- Em cada célula de resolução, assume-se a presença de um grande número de espalhadores.
- Pelo Teorema Central do Limite, as partes real $R[n, m]$ e imaginária $C[n, m]$ do sinal complexo recebido são independentes e distribuem-se aproximadamente como:

$$R[n, m], C[n, m] \sim \mathcal{N}(0, \sigma^2).$$

- Definição do pixel SAR:

$$S[n, m] = R[n, m] + iC[n, m], \quad i = \sqrt{-1}.$$

A amplitude é:

$$Y[n, m] = |S[n, m]| = \sqrt{R[n, m]^2 + C[n, m]^2},$$

que segue exatamente uma distribuição de Rayleigh:

$$f(y; \sigma) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right), \quad y > 0.$$

Por que a Distribuição de Rayleigh em SAR?

- Relevância prática:
 - ▶ Modelo amplamente aceito pela comunidade de sensoriamento remoto para regiões homogêneas.
 - ▶ Base para modelos mais gerais (K , \mathcal{G}^0) usados em regiões heterogêneas.
 - ▶ Amplitude resultante de um grande número de espalhadores iid.

Aplicações de Sensoriamento Remoto

- Monitoramento de uso e cobertura da terra.
- Detecção de desmatamento, queimadas e mudanças no ambiente.
- Aplicações agrícolas: estimativa de produtividade, irrigação.
- Monitoramento de desastres: enchentes, deslizamentos, manchas de óleo.

Motivação e Contexto

Modelagem de imagens SAR

- Identificação e classificação de alvos ou usos do solo distintos nas imagens;
- Esses problemas são geralmente tratados assumindo **homogeneidade** nas regiões;
- Homogeneidade: parâmetros constantes;
- Modelo de regressão:
 - ▶ Não assume homogeneidade nas imagens;
 - ▶ Permite tratar os desafios específicos de imagens SAR.

Motivação e Contexto

Tópicos a serem abordados

- Inferências dos parâmetros do modelo de regressão são baseadas nos estimadores de máxima verossimilhança (MLE);
- MLEs possuem boas propriedades assintóticas;
- Porém, em muitos casos temos comprimentos de sinal (ou número de pixels) pequenos;
- Necessidade de correções inferenciais para amostras pequenas.
- Estatísticas afetadas pela presença de outliers.

Motivação e Contexto

Modelagem de imagens SAR

- Pixels podem apresentar diferentes níveis de dependência:
 - ▶ Dependência espacial (vizinhança dos pixels);
 - ▶ Dependência temporal (imagens em série temporal).
- Um modelo ARMA bidimensional (2D) é uma alternativa para capturar essas dependências.

Detector de Mudanças — Distribuição Rayleigh

Detecção de Mudanças em Imagens SAR

O que é Detecção de Mudanças?

- Processo de identificar diferenças significativas entre duas ou mais imagens da mesma região, adquiridas em tempos distintos.
- Utilizada para detectar eventos como:
 - ▶ desmatamento,
 - ▶ expansão urbana,
 - ▶ inundações,
 - ▶ surgimento de alvos.
- Métodos podem ser baseados, por exemplo, em:
 - ▶ testes de hipótese,
 - ▶ gráficos de controle,
 - ▶ entropia de Shannon.

Detector de Mudanças — Distribuição Rayleigh

O que é Entropia de Shannon?

- Mede a incerteza ou o conteúdo de informação dos dados — quanto maior a entropia, maior a aleatoriedade.

$$H_S = - \int_0^{\infty} f(y) \log f(y) dy$$

Detector de Mudanças — Distribuição Rayleigh

Etapas do Método:

- ① Modelagem dos Pixels: Cada pixel é modelado pela distribuição de Rayleigh, adequada para dados SAR homogêneos.
- ② Cálculo da Entropia: Para cada janela local (ex.: 3×3), calcula-se a entropia de Shannon a partir da estimativa de máxima verossimilhança de σ .
- ③ Teste de Hipótese:
 H_0 : Entropia (Imagem 1) = Entropia (Imagem 2)
 H_1 : Entropias diferentes \Rightarrow Mudança detectada.

- ④ Janela Deslizante e Decisão: O teste é aplicado pixel a pixel sobre toda a imagem.

Detector de Mudanças — Distribuição Rayleigh

Entropia (Rayleigh):

$$H_S(\beta) = 1 + \ln\left(\frac{\beta}{\sqrt{2}}\right) + \frac{\gamma}{2}, \quad \hat{\beta} = \sqrt{\frac{1}{2N} \sum_{n=1}^N y_n^2}, \quad \hat{H}_S = H_S(\hat{\beta}).$$

Variância assintótica de \hat{H}_S (Lema):

$$\delta = \frac{\partial H_S}{\partial \beta} = \frac{1}{\beta}, \quad \mathcal{K}(\beta) = \frac{4N}{\beta^2} \Rightarrow \text{Var}(\hat{H}_S) \approx \delta^\top \mathcal{K}(\beta)^{-1} \delta = \frac{1}{4N}.$$

Detector de Mudanças — Distribuição Rayleigh

Teste entre duas janelas (amostras) de tamanhos N_1, N_2 (Teorema):

$$S_S = \sum_{i=1}^2 \frac{N_i (\hat{H}_{S,i} - \bar{\nu})^2}{\sigma_i^2}, \quad \bar{\nu} = \frac{\sum_{i=1}^2 \frac{N_i \hat{H}_{S,i}}{\sigma_i^2}}{\sum_{i=1}^2 \frac{N_i}{\sigma_i^2}}, \quad \sigma_i^2 \approx \frac{1}{4},$$

Sob H_0 : $S_S \xrightarrow{d} \chi_1^2$.

Detector de Mudanças — Distribuição Rayleigh

Teste entre duas janelas (amostras) de tamanhos N_1, N_2 (Teorema):

$$S_S = \sum_{i=1}^2 \frac{N_i (\hat{H}_{S,i} - \bar{\nu})^2}{\sigma_i^2}, \quad \bar{\nu} = \frac{\sum_{i=1}^2 \frac{N_i \hat{H}_{S,i}}{\sigma_i^2}}{\sum_{i=1}^2 \frac{N_i}{\sigma_i^2}}, \quad \sigma_i^2 \approx \frac{1}{4},$$

$$\text{Sob } H_0 : S_S \xrightarrow{d} \chi_1^2.$$

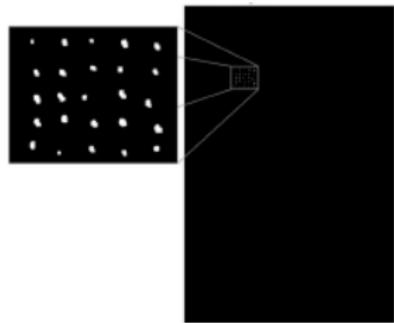
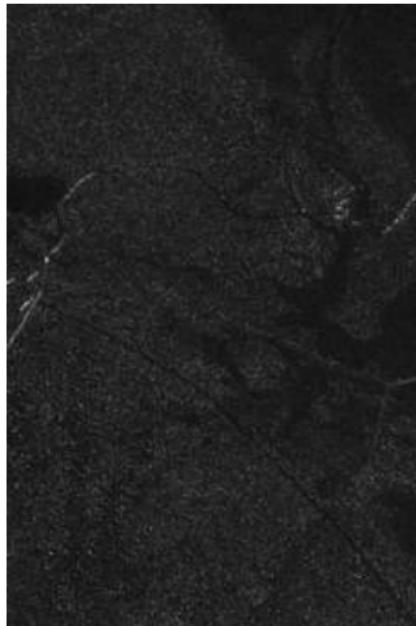
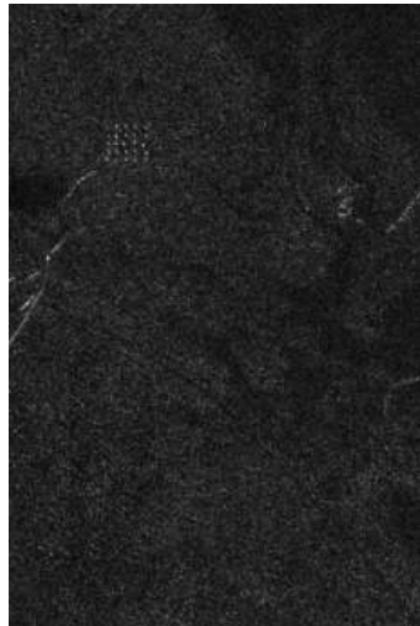
Caso balanceado ($N_1 = N_2 = N$) (Proposição):

$$S_S = N \frac{(\hat{H}_{S,1} - \hat{H}_{S,2})^2}{\sigma_1^2 + \sigma_2^2} = 2N (\hat{H}_{S,1} - \hat{H}_{S,2})^2.$$

Regra de decisão:

$$p\text{-valor} = 1 - F_{\chi_1^2}(S_S) \Rightarrow \text{rejeita } H_0 \text{ se } p \leq \alpha.$$

Detector de Mudanças — Distribuição Rayleigh — Análise com Dados Reais



Detector de Mudanças — Distribuição Rayleigh — Análise com Dados Reais



2009



2015



Modelo de Regressão Rayleigh

A Distribuição de Rayleigh

Seja Y uma variável aleatória com distribuição de Rayleigh com parâmetro de média $\mu > 0$. A função densidade de probabilidade da distribuição de Rayleigh baseada na média é dada por

$$f_Y(y; \mu) = \frac{\pi y}{2\mu^2} \exp\left(-\frac{\pi y^2}{4\mu^2}\right), \quad y > 0.$$

A função de distribuição acumulada é dada por

$$F_Y(y; \mu) = 1 - \exp\left(-\frac{\pi y^2}{4\mu^2}\right).$$

Modelo de Regressão de Rayleigh

Suposições

- $Y[1], Y[2], \dots, Y[N]$ são N amostras aleatórias independentes;
- Cada amostra segue a densidade de Rayleigh com média $\mu[n]$;
- Cada amostra assume valores $y[n], n = 1, 2, \dots, N$.

Modelo de Regressão de Rayleigh

O modelo

$$\eta[n] = g(\mu[n]) = \sum_{i=1}^r \beta_i x_i[n], \quad n = 1, 2, \dots, N,$$

em que:

- $r < N$ é o número de covariáveis consideradas no modelo;
- $\beta = (\beta_1, \beta_2, \dots, \beta_r)^\top$ é um vetor de parâmetros lineares desconhecidos;
- $x[n] = (x_1[n], x_2[n], \dots, x_r[n])^\top$ é um vetor de variáveis independentes determinísticas;
- $g(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável, onde $g : \mathbb{R}^+ \rightarrow \mathbb{R}$.

Processo de Estimação

Os estimadores de máxima verossimilhança são dados por

$$\hat{\beta} = \arg \max_{\beta} \ell(\beta),$$

em que $\ell(\beta)$ é a função de log-verossimilhança dos parâmetros para o sinal observado, definida como

$$\ell(\beta) = \sum_{n=1}^N \ell[n](\mu[n]).$$

Processo de Estimação

A quantidade $\ell[n](\mu[n])$ é o logaritmo de $f(y[n], \mu[n])$, que pode ser escrito como

$$\ell[n](\mu[n]) = \log\left(\frac{\pi}{2}\right) + \log(y[n]) - \log(\mu[n]^2) - \frac{\pi y[n]^2}{4\mu[n]^2},$$

em que $\mu[n] = g^{-1} \left(\sum_{i=1}^k x_i[n] \beta_i \right)$.

Processo de Estimação

O vetor escore, obtido ao diferenciar a log-verossimilhança em relação a cada parâmetro desconhecido β_i , $i = 1, 2, \dots, k$, é dado por

$$U(\boldsymbol{\beta}) = \left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_1}, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_2}, \dots, \frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_k} \right)^\top.$$

Processo de Estimação

Considerando a regra da cadeia, temos que

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{n=1}^N \frac{d\ell[n](\mu[n])}{d\mu[n]} \frac{d\mu[n]}{d\eta[n]} \frac{\partial \eta[n]}{\partial \beta},$$

em que

$$\frac{d\ell[n](\mu[n])}{d\mu[n]} = \frac{\pi y[n]^2}{2\mu[n]^3} - \frac{2}{\mu[n]},$$

$$\frac{d\mu[n]}{d\eta[n]} = \frac{1}{g'(\mu[n])},$$

$$\frac{\partial \eta[n]}{\partial \beta_i} = x_i[n],$$

e $g'(\cdot)$ é a primeira derivada da função de ligação adotada $g(\cdot)$.

Processo de Estimação

Na forma matricial, o vetor escore pode ser escrito como

$$U(\beta) = X^\top \cdot T \cdot v,$$

em que X é uma matriz $N \times k$ cuja n -ésima linha é $x[n]^\top$, e

$$\begin{aligned} T &= \text{diag} \left\{ \frac{1}{g'(\mu[1])}, \frac{1}{g'(\mu[2])}, \dots, \frac{1}{g'(\mu[N])} \right\}, \\ v &= \left(\frac{\pi y[1]^2}{2\mu[1]^3} - \frac{2}{\mu[1]}, \dots, \frac{\pi y[N]^2}{2\mu[N]^3} - \frac{2}{\mu[N]} \right)^\top. \end{aligned}$$

Processo de Estimação

Assim, o estimador de máxima verossimilhança (EMV) para cada parâmetro do modelo de regressão de Rayleigh é obtido resolvendo o seguinte sistema não linear

$$U(\beta) = 0,$$

em que 0 é o vetor de zeros de dimensão k .

Para isso, o método BFGS com primeiras derivadas analíticas foi considerado.

Inferência de Teste

A matriz de informação de Fisher é dada por

$$I(\beta) = X^\top \cdot W \cdot X,$$

em que

$$W = \text{diag} \left\{ \frac{4}{\mu[1]^2} \left(\frac{d\mu[1]}{d\eta[1]} \right)^2, \dots, \frac{4}{\mu[N]^2} \left(\frac{d\mu[N]}{d\eta[N]} \right)^2 \right\}.$$

Modelo de Regressão de Rayleigh - Teoria de Detecção

- O problema de detecção consiste em distinguir entre as hipóteses:

$$\begin{cases} \mathcal{H}_0 : \mu[n] = g^{-1}(\beta_1 + \sum_{i=3}^r \beta_i x_i[n]), & (\beta_2 = 0), \\ \mathcal{H}_1 : \mu[n] = g^{-1}(\beta_1 + \beta_2 x_2[n] + \sum_{i=3}^r \beta_i x_i[n]). \end{cases}$$

- Na situação em que $\beta_2 \neq 0$, o uso do solo como floresta é detectado.

Modelo de Regressão de Rayleigh - Teoria de Detecção

- Para derivar o detector, podemos usar o teste de Wald, definido como:

$$T_W = (\hat{\beta}_{I1} - \beta_{I0})^\top \left(\left[I^{-1}(\hat{\beta}_1) \right]_{\beta_I \beta_I} \right)^{-1} (\hat{\beta}_{I1} - \beta_{I0}),$$

em que:

- $\hat{\beta}_1 = (\hat{\beta}_{I1}^\top, \hat{\beta}_{M1}^\top)^\top$ é o estimador de máxima verossimilhança sob $\beta_I \neq \beta_{I0}$;
- β_I é o vetor de parâmetros de interesse;
- β_M é o vetor de parâmetros de incômodo;
- $\left[I^{-1}(\hat{\beta}) \right]_{\beta_I \beta_I}$ é a partição de $I(\hat{\beta})$ limitada às estimativas de interesse.

Modelo de Regressão de Rayleigh - Análise com Dados Reais

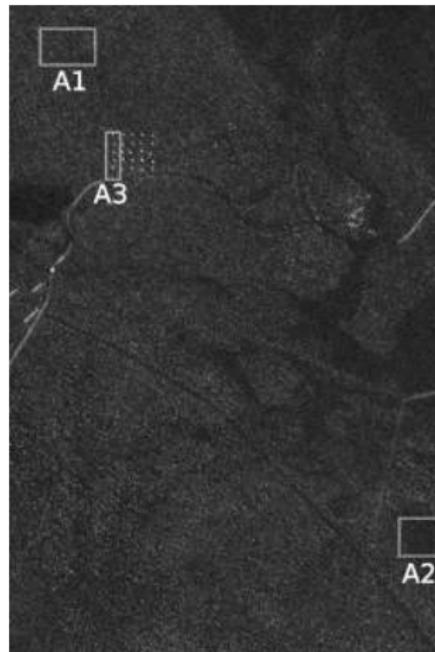


Figura: Imagem SAR CARABAS II mostrando as regiões testadas.

Modelo de Regressão de Rayleigh - Análise com Dados Reais

- Testamos três regiões de uma imagem SAR;
- O modelo é especificado como:

$$g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n],$$

em que:

- ▶ $x_1[n] = 1$ é o intercepto;
- ▶ $x_2[n]$ e $x_3[n]$ são duas variáveis dummy representando cada região testada.

Modelo de Regressão de Rayleigh - Análise com Dados Reais

Neste caso, temos os seguintes quatro modelos:

- ① $g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n]$
- ② $g(\mu[n]) = \beta_1$, se $\beta_2[n] = \beta_3[n] = 0$
- ③ $g(\mu[n]) = \beta_1 + \beta_2 x_2[n]$, se $\beta_3[n] = 0$
- ④ $g(\mu[n]) = \beta_1 + \beta_3 x_3[n]$, se $\beta_2[n] = 0$

Modelo de Regressão de Rayleigh - Análise com Dados Reais

Tabela: Modelos de regressão ajustados para as Regiões A1, A2, e A3

Estimativa	Erro Padrão	Detecção (<i>p</i> -valor)
Modelo de regressão de Rayleigh		
$\hat{\beta}_1$	−2.0623	0.0445
$\hat{\beta}_2$	−0.1280	0.0599
$\hat{\beta}_3$	1.0784	0.0616
Modelo de regressão Gaussiano		
$\hat{\beta}_1$	0.12683	0.01646
$\hat{\beta}_2$	−0.01201	0.02213
$\hat{\beta}_3$	0.15948	0.02277
GLM Gama		
$\hat{\beta}_1$	7.8844	0.5209
$\hat{\beta}_2$	0.8248	0.7341
$\hat{\beta}_3$	−4.3917	0.5657

Correção de Viés no Modelo de Regressão Rayleigh

Motivação e Objetivo

- Problema: EMV pode ser enviesado para janelas pequenas (e.g., 3×3).

Meta

Propor e avaliar estimadores com correção de viés para o Modelo de Regressão de Rayleigh quando o comprimento do sinal (N) é pequeno.

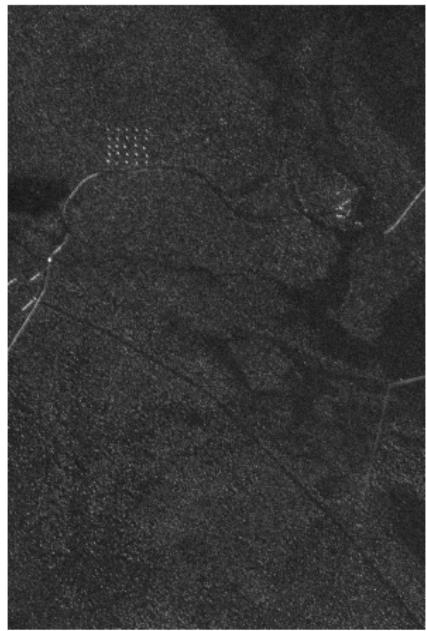
- Métodos: Cox–Snell, Firth e Bootstrap Paramétrico.
- Simulações: Firth frequentemente apresenta a menor norma quadrática integrada do viés relativo (IRBSN) entre os métodos comparados.

Experimentos com dados SAR

Imagens: San Francisco Bay (canais HH, VV, HV) e CARABAS II (UWB VHF).

- Modelos HH~HV e VV~HV (HV como covariável para a média dos outros canais)
— SF.
- Outra imagem do mesmo pacote como covariável (imagens correlacionadas) — CARABAS.
- Janelas de 3×3 e 7×7 na região de água.

Experimentos com dados SAR



(a) CARABAS II



(b) San Francisco

Figura: San Francisco and CARABAS II SAR images associated to HH polarization channel.

Experimentos com dados SAR

Tabela: Modelos ajustados para imagens SAR de San Francisco e CARABAS II

		MLE		Firth	
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$
SF — HH ~ HV	3 × 3	-2.9162	0.3091	-2.9318	3.6062
	7 × 7	-2.8978	0.0643	-2.9115	1.2344
SF — VV ~ HV	3 × 3	-2.2259	-0.1947	-2.1421	-1.3629
	7 × 7	-2.1621	0.0101	-2.1559	0.1569
CARABAS II	3 × 3	-1.9124	-0.4279	-1.7851	-1.2194
	7 × 7	-1.9896	0.1616	-1.9886	0.2314

Experimentos com dados SAR

Tabela: Erro quadrático médio estimado dos modelos ajustados às imagens SAR de San Francisco e do CARABAS II

	SF — HH ~ HV	Baseado em Firth		Baseado em MV	
		Rayleigh regressão	Rayleigh regressão	Gaussiana regressão	Rayleigh distribuição
SF — HH ~ HV	3 × 3	0.0123	0.0133	1.0019	0.5662
	7 × 7	0.0157	0.0159	1.0020	0.5606
SF — VV ~ HV	3 × 3	0.0194	0.0222	1.0077	0.5085
	7 × 7	0.0326	0.0330	1.0088	0.4955
CARABAS II	3 × 3	0.0267	0.0296	1.0136	0.4730
	7 × 7	0.0427	0.0430	1.0130	0.4774

Estimadores Robustos no Modelo de Regressão Rayleigh

Modelo de Regressão Rayleigh

Para o modelo derivado, a estimativa baseada no processo de máxima verossimilhança foi considerada.

Problema!

Tal método de inferência é assintoticamente eficiente, mas carece de robustez contra misspecificação do modelo e outliers.

Proposta

Utilizar o método de máxima verossimilhança ponderada para estimar os parâmetros do modelo de regressão Rayleigh, visando obter estimadores robustos à presença de outliers.

Processo de Estimação

As estimativas de máxima verossimilhança ponderada são dadas por

$$\hat{\beta} = \arg \max_{\beta} \ell_w(\beta),$$

em que $\ell_w(\beta)$ é a log-verossimilhança ponderada dos parâmetros para o sinal observado, definida como

$$\ell_w(\beta) = \sum_{n=1}^N w[n] \ell[n](\mu[n]).$$

Processo de Estimação

O vetor de pesos $w[n]$ é determinado como:

$$w[n] = \begin{cases} \frac{F(y[n]; \mu[n])}{\delta}, & \text{se } F(y[n]; \mu[n]) < \delta, \\ 1, & \text{se } \delta \leq F(y[n]; \mu[n]) \leq 1 - \delta, \\ \frac{1 - F(y[n]; \mu[n])}{\delta}, & \text{se } F(y[n]; \mu[n]) > 1 - \delta, \end{cases}$$

em que $\delta \in (0, 1)$.

Processo de Estimação

O vetor de pesos $w[n]$ é determinado como:

$$w[n] = \begin{cases} \frac{F(y[n]; \mu[n])}{\delta}, & \text{se } F(y[n]; \mu[n]) < \delta, \\ 1, & \text{se } \delta \leq F(y[n]; \mu[n]) \leq 1 - \delta, \\ \frac{1 - F(y[n]; \mu[n])}{\delta}, & \text{se } F(y[n]; \mu[n]) > 1 - \delta, \end{cases}$$

em que $\delta \in (0, 1)$.

Valores atípicos de $y[n]$: $\uparrow\downarrow F(y[n]; \mu[n])$ valores → ponderados → **distorções de inferência minimizadas**.

Simulação de Monte Carlo

ϵ	Measures	WMLE			MLE		
		$\hat{\beta}_1$	$\hat{\beta}_2$	Absolute Total	$\hat{\beta}_1$	$\hat{\beta}_2$	Absolute Total
$N = 100$							
0%	Mean	0.4957	0.1488	—	0.4967	0.1486	—
	RB(%)	-0.8570	-0.8259	1.6828	-0.6598	-0.9167	1.5765
	MSE	0.0100	0.0312	0.0413	0.0099	0.0309	0.0409
1%	Mean	0.4968	0.1480	—	0.6401	0.1247	—
	RB(%)	-0.6311	-1.3311	1.9622	28.0262	-16.8944	44.9206
	MSE	0.0101	0.0316	0.0417	0.0715	0.1916	0.2631
5%	Mean	0.5476	0.1597	—	1.0103	0.0705	—
	RB(%)	9.5122	6.4767	15.9889	102.0672	-53.0026	155.0698
	MSE	0.0518	0.1693	0.2211	0.3405	0.3098	0.6503
$N = 750$							
0%	Mean	0.4980	0.1507	—	0.4996	0.1506	—
	RB(%)	-0.3962	0.4568	0.8530	-0.0875	0.4140	0.5015
	MSE	0.0014	0.0041	0.0054	0.0014	0.0040	0.0054
1%	Mean	0.4997	0.1507	—	0.6673	0.1184	—
	RB(%)	-0.0593	0.4608	0.5201	33.4529	-21.0987	54.5516
	MSE	0.0014	0.0041	0.0054	0.0360	0.0300	0.0660
5%	Mean	0.5080	0.1631	—	1.0359	0.0698	—
	RB(%)	1.6012	8.7438	10.3450	107.1780	-53.4431	160.6211
	MSE	0.0027	0.0099	0.0127	0.2959	0.0396	0.3355

Modelagem de Imagens SAR - Detector de Mudança de Tipo de Solo

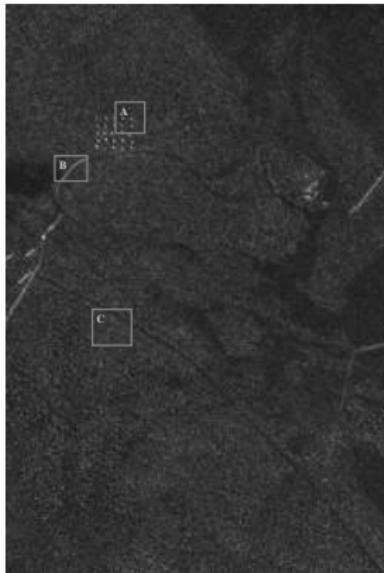


Figura: Imagem CARABAS II — As Regiões A, B e C representam, respectivamente, áreas contendo veículos militares, linhas de transmissão e floresta.

Modelagem de Imagens SAR - Detector de Mudança de Tipo de Solo

A detecção de uma área de tipo de solo de interesse pode ser obtida por:

$$g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n],$$

em que

- β_1 é o intercepto;
- $x_2[n]$ e $x_3[n]$ são covariáveis binárias iguais a um se a região consiste na região de interesse e zero caso contrário.

Modelagem de Imagens SAR - Detector de Mudança de Tipo de Solo

Tabela: Modelos de regressão ajustados para as Regiões A, B e C

	Estimativa	Erro Padrão	p-valor
Modelo de regressão Rayleigh (estimadores robustos)			
$\hat{\beta}_1$	-1.6681	0.0333	< 0.001
$\hat{\beta}_2$	0.1168	0.0521	0.0250
$\hat{\beta}_3$	-0.4993	0.0521	< 0.001
Modelo de regressão Rayleigh (estimadores não robustos)			
$\hat{\beta}_1$	-1.4916	0.0333	< 0.001
$\hat{\beta}_2$	-0.0555	0.0521	0.2867
$\hat{\beta}_3$	-0.6759	0.0521	< 0.001
Modelo de regressão Gaussiano (estimadores robustos)			
$\hat{\beta}_1$	0.1896	0.0090	< 0.001
$\hat{\beta}_2$	0.0098	0.0140	0.4870
$\hat{\beta}_3$	-0.0738	0.0140	< 0.001
Modelo de regressão Gama (estimadores robustos)			
$\hat{\beta}_1$	5.5817	0.2558	< 0.001
$\hat{\beta}_2$	-0.4396	0.3823	0.2500
$\hat{\beta}_3$	2.6408	0.5209	< 0.001

Modelagem de Imagens SAR - Detector de Mudança de Tipo de Solo



Figura: Imagem OrbiSAR — As Regiões D, E e F representam, respectivamente, uma área aberta, uma estrada e uma área urbana.

Modelagem de Imagens SAR - Detector de Mudança de Tipo de Solo

Tabela: Modelos de regressão ajustados para as Regiões D, E e F

	Estimativa	Erro Padrão	p-valor
Modelo de regressão Rayleigh (estimadores robustos)			
$\hat{\beta}_1$	-0.6069	0.0423	< 0.001
$\hat{\beta}_2$	-0.1190	0.0598	0.0464
$\hat{\beta}_3$	2.7767	0.1319	< 0.001
Modelo de regressão Rayleigh (estimadores não robustos)			
$\hat{\beta}_1$	-0.6069	0.0423	< 0.001
$\hat{\beta}_2$	-0.1105	0.0598	0.0646
$\hat{\beta}_3$	2.6749	0.1319	< 0.001
Modelo de regressão Gaussiano (estimadores robustos)			
$\hat{\beta}_1$	0.6087	0.1038	< 0.001
$\hat{\beta}_2$	-0.0777	0.1468	0.5970
$\hat{\beta}_3$	6.6209	0.3241	< 0.001

Modelagem de Imagens SAR - Detecção de Anomalias

Proposta do Detector

Propomos uma metodologia de detecção de anomalias baseada no resíduo do modelo de regressão Rayleigh.

- Os resíduos quantílicos são definidos como:

$$r[n, m] = \Phi^{-1}(F_Y(y[n, m] | S[n, m])),$$

onde Φ^{-1} é a função quantílica da normal padrão;

- Quando o modelo está corretamente ajustado: $r[n, m] \sim \mathcal{N}(0, 1)$.

Modelagem de Imagens SAR - Detecção de Anomalias

Premissas do Detector

Modelo ajustado corretamente →

Modelagem de Imagens SAR - Detecção de Anomalias

Premissas do Detector

Modelo ajustado corretamente → os resíduos são distribuídos aleatoriamente em torno de zero e dentro do intervalo $[-3, 3]$.

Modelagem de Imagens SAR - Detecção de Anomalias

Premissas do Detector

Modelo ajustado corretamente \rightarrow os resíduos são distribuídos aleatoriamente em torno de zero e dentro do intervalo $[-3, 3]$.

Valor do resíduo fora do intervalo $[-3, 3] \rightarrow$

Modelagem de Imagens SAR - Detecção de Anomalias

Premissas do Detector

Modelo ajustado corretamente \rightarrow os resíduos são distribuídos aleatoriamente em torno de zero e dentro do intervalo $[-3, 3]$.

Valor do resíduo fora do intervalo $[-3, 3]$ \rightarrow o pixel analisado é entendido como diferente do comportamento esperado da região de interesse \rightarrow

Modelagem de Imagens SAR - Detecção de Anomalias

Premissas do Detector

Modelo ajustado corretamente \rightarrow os resíduos são distribuídos aleatoriamente em torno de zero e dentro do intervalo $[-3, 3]$.

Valor do resíduo fora do intervalo $[-3, 3]$ \rightarrow o pixel analisado é entendido como diferente do comportamento esperado da região de interesse \rightarrow **uma anomalia é detectada.**

Modelagem de Imagens SAR - Detecção de Anomalias

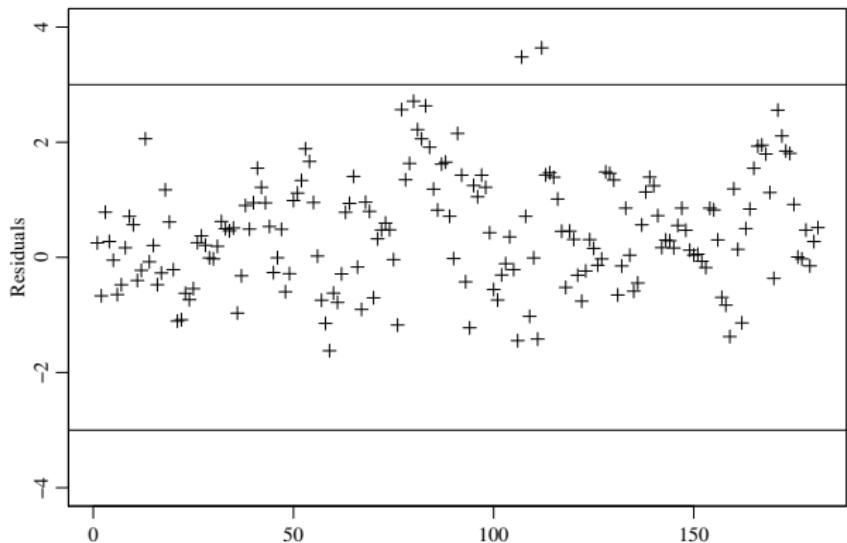


Figura: Gráfico de controle de resíduos (1-D - motivação).

Modelagem de Imagens SAR - Detecção de Anomalias

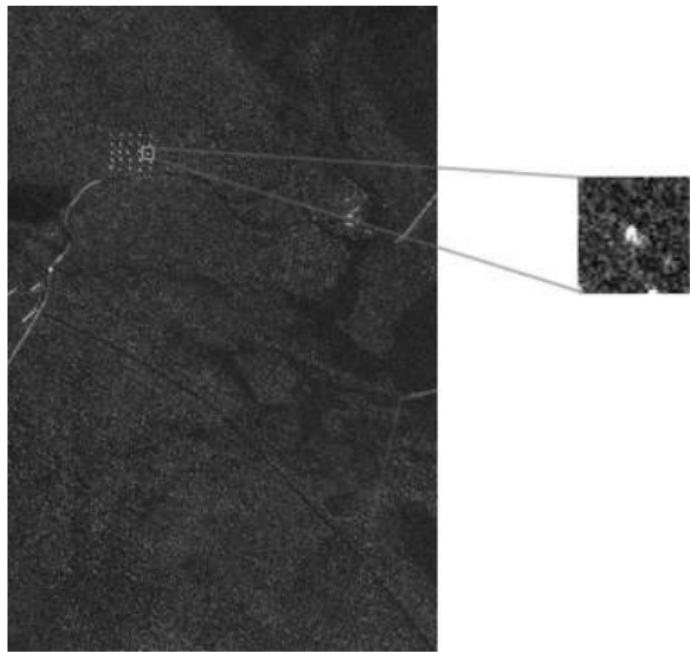


Figura: Imagem SAR avaliada e a região de interesse usada para realizar a detecção de anomalias.

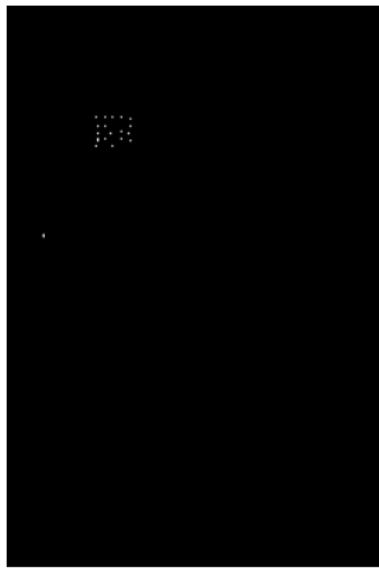
Modelagem de Imagens SAR - Detecção de Anomalias

Para ajustar o modelo de regressão, consideramos a seguinte estrutura:

$$g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n] + \beta_4 x_4[n],$$

em que $x_2[n]$, $x_3[n]$, e $x_4[n]$ foram extraídas de as outras três imagens com o mesmo voo disponíveis no conjunto de dados CARABAS II.

Modelagem de Imagens SAR - Detecção de Anomalias



(a) Estimação baseada em robustez



(b) Estimação não robusta

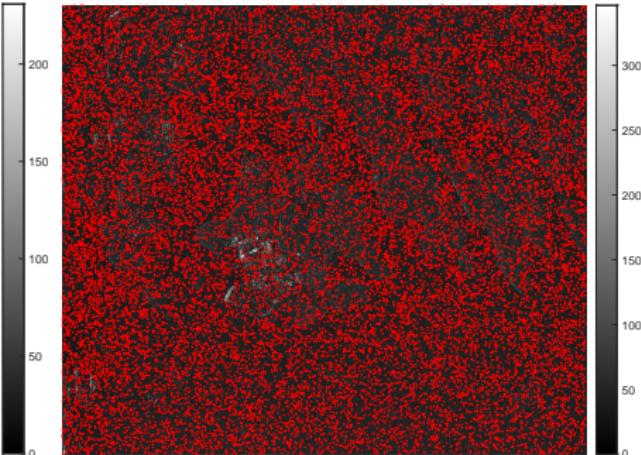
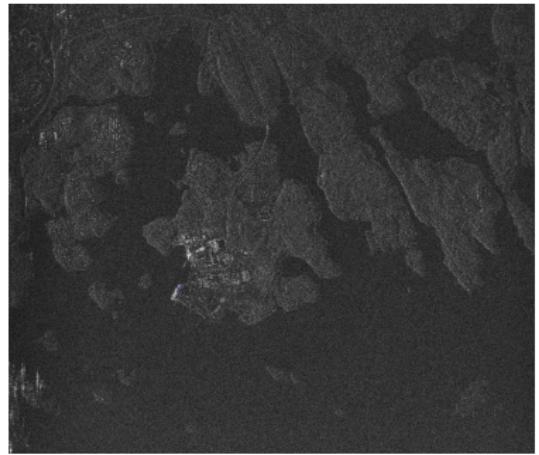
Figura: Resultados de detecção baseados no modelo de regressão Rayleigh robusto e não robusto.

Distribuição Rayleigh Inflacionada

Motivação

Valores Zero em Imagens SAR:

- Grandes variações de intensidade podem resultar em pixels pretos, levando a cenários de modelagem não realistas.
- Valores nulos de amplitude ou intensidade podem estar distribuídos aleatoriamente, exigindo modelos estatísticos que incorporem esses zeros sem distorcer as estatísticas gerais da imagem.
- A distribuição Rayleigh inflacionada foi introduzida para modelar tanto os valores contínuos quanto os pixels com valor zero.



Distribuição Rayleigh Inflacionada

- Trata de casos em que os sinais contêm zeros.
- Mistura da distribuição Rayleigh com a distribuição degenerada em zero:

$$\bar{F}_Y(y; \lambda, \mu) = \lambda + (1 - \lambda)F_Y(y; \mu), \quad 0 < \lambda < 1.$$

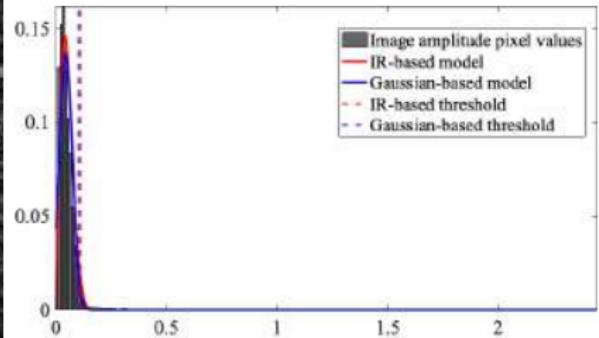
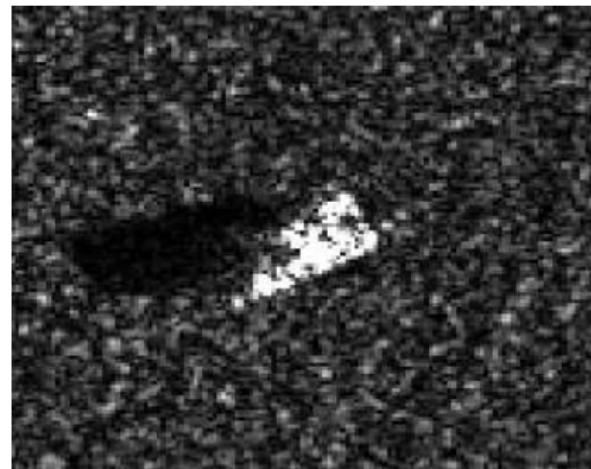
- Função densidade de probabilidade correspondente (PDF):

$$\bar{f}_Y(y; \lambda, \mu) = \begin{cases} \lambda, & y = 0, \\ (1 - \lambda)f_Y(y; \mu), & y > 0. \end{cases}$$

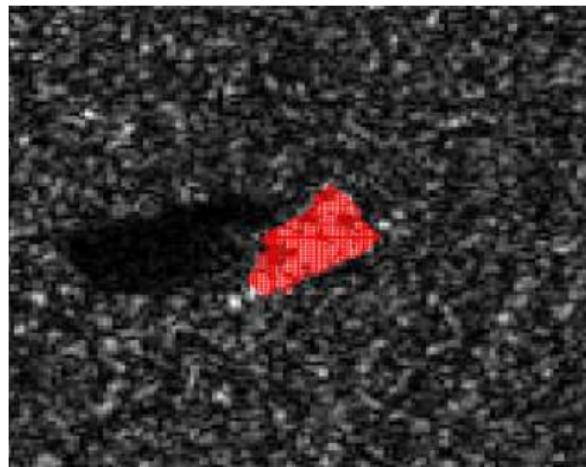
- Média e variância:

$$\mathbb{E}(Y) = (1 - \lambda)\mu, \quad \text{Var}(Y) = \mu^2(1 - \lambda) \left[\frac{4}{\pi} - 2 + \lambda \right].$$

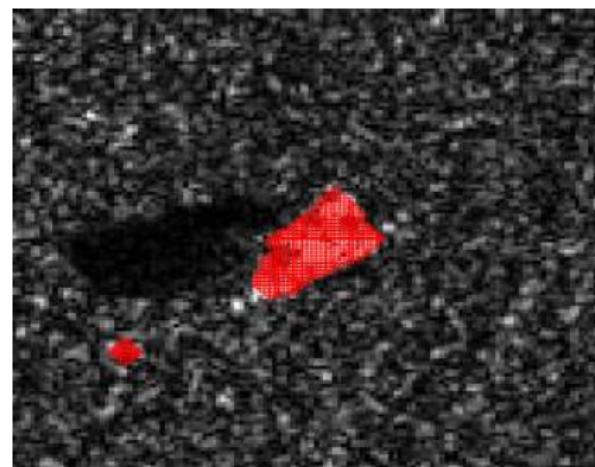
Dados MSTAR



Dados MSTAR

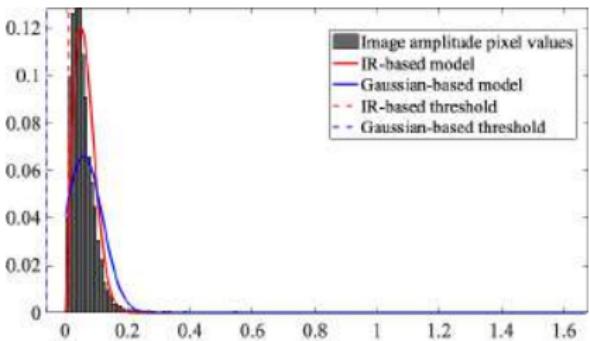
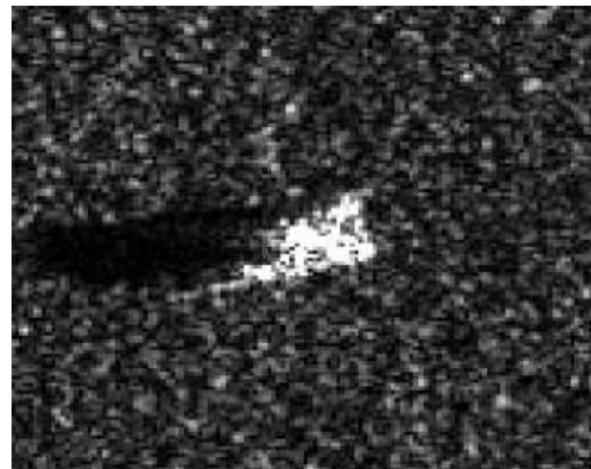


Inflated Rayleigh distribution

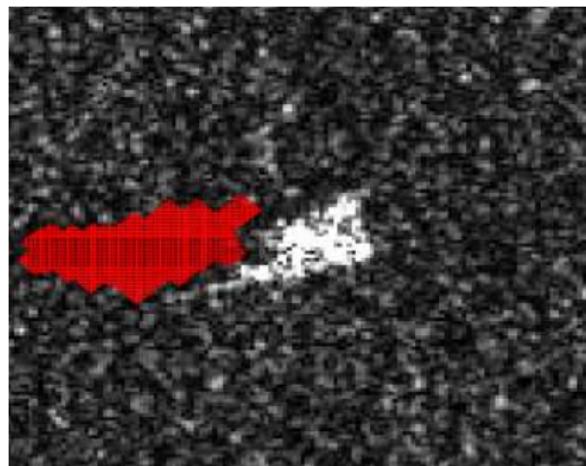


Gaussian distribution

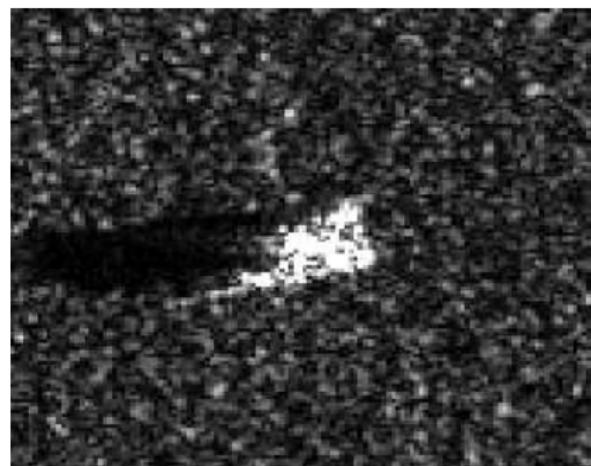
Dados MSTAR



Dados MSTAR



Inflated Rayleigh distribution



Gaussian distribution

Modelo de Regressão Rayleigh Inflacionado

Modelo de Regressão

- Modelo de regressão Rayleigh inflacionado com preditores lineares:

$$\eta_1[n] = g_1(\lambda[n]) = \sum_{j=1}^m z_j[n] \gamma_j,$$

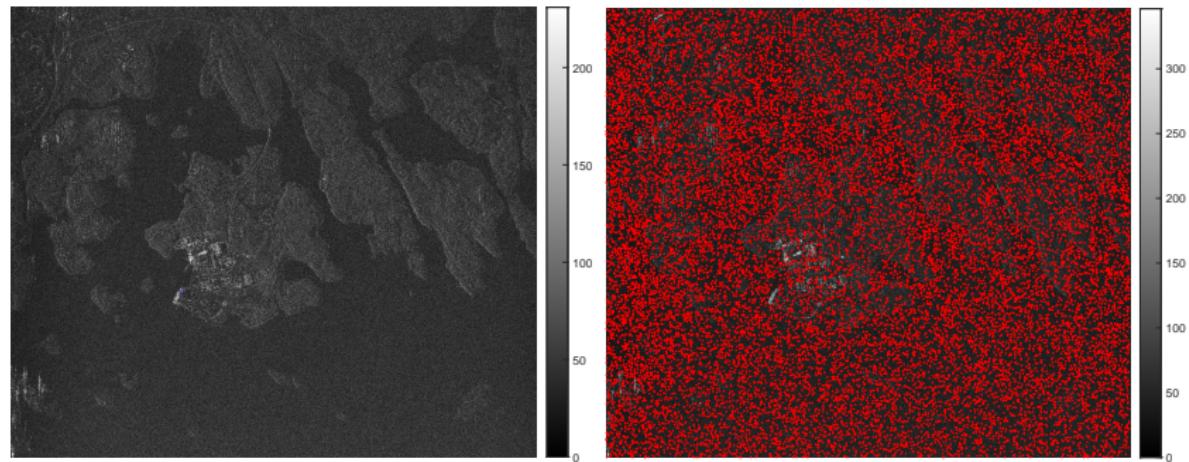
$$\eta_2[n] = g_2(\mu[n]) = \sum_{i=1}^r x_i[n] \beta_i.$$

- Funções de ligação:
 - ▶ g_1 : logito, probito, log-log, log-log complementar, Cauchy.
 - ▶ g_2 : logaritmo ou raiz quadrada.
- Se η_1 contém apenas um intercepto, então λ é a proporção de zeros.

Análise com Dados Reais

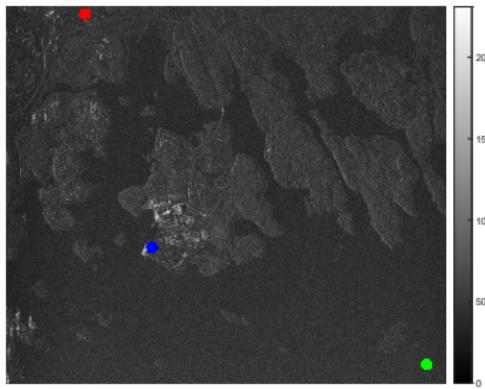
Imagen SAR:

- Fonte: ICEYE Radar, Karlskrona, Suécia.
- A imagem SAR testada possui 25728×7808 pixels e 22319 (aproximadamente 1% do total) têm valor de pixel nulo.

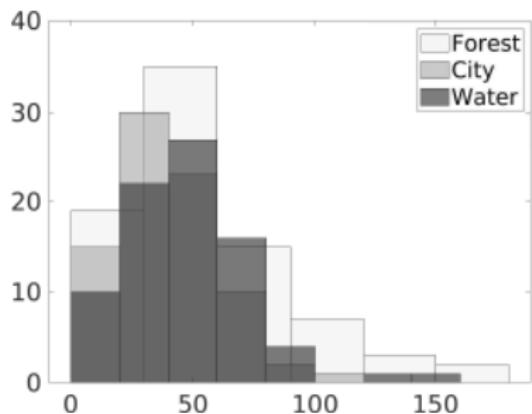


Metodologia de Detecção

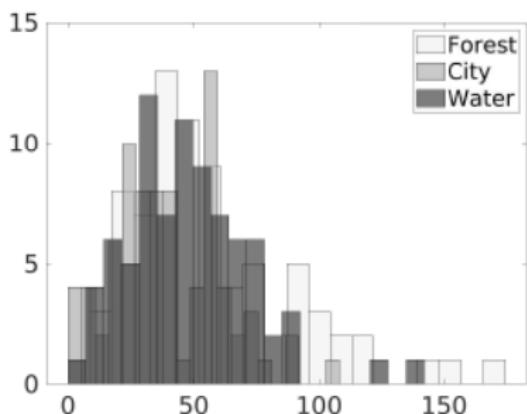
- O objetivo é distinguir entre três regiões em uma imagem SAR da ICEYE, correspondentes a diferentes tipos de solo: floresta, cidade e mar.
- Foi utilizada uma janela de 4×4 pixels para a detecção, com um total de aproximadamente 250 observações por região.



Metodologia de Detecção



(a) 6 classes



(b) 20 classes

Figura: Histogramas dos três diferentes tipos de terreno avaliados.

Metodologia de Detecção

- O sinal de resposta foi modelado utilizando a seguinte estrutura:

$$g_1(\lambda[n]) = \gamma_1,$$

$$g_2(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n].$$

- Na equação, $y[n]$ é a magnitude vetorizada dos pixels, e $x_2[n]$ e $x_3[n]$ representam os tipos de terreno urbano e mar, respectivamente.
- O experimento estima λ considerando apenas o intercepto, assumindo que os valores de pixels nulos são independentes do tipo de terreno.

Metodologia de Detecção

Tabela: Estimativa, erro-padrão e *p*-valor dos modelos de regressão ajustados para as regiões testadas

	Estimativa	Erro-Padrão	<i>p</i> -valor
Modelo de regressão Rayleigh Inflacionado			
$\hat{\gamma}_1$	-4.3821	0.5810	< 0.0001
$\hat{\beta}_1$	4.0505	0.0556	< 0.0001
$\hat{\beta}_2$	-0.3438	0.0786	< 0.0001
$\hat{\beta}_3$	-0.1784	0.0786	0.0232
Modelo de regressão Gaussiano			
$\hat{\beta}_1$	54.62	3.0725	< 0.0001
$\hat{\beta}_2$	-14.748	4.3452	< 0.0001
$\hat{\beta}_3$	-6.6979	4.3452	0.12453

iRSARMAX(p, q) \times (P, Q)_S: Estrutura

$$g(\lambda_t) = \alpha_0 + \alpha_1 Y_{t-1},$$

$$\begin{aligned} h(\mu_t) &= \beta_0 + \mathbf{x}_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^P \Phi_j Y_{t-jS} \\ &\quad - \sum_{i=1}^p \sum_{j=1}^P \phi_i \Phi_j Y_{t-(i+jS)} - \sum_{a=1}^q \theta_a r_{t-a} - \sum_{b=1}^Q \Theta_b r_{t-bS} \\ &\quad + \sum_{a=1}^q \sum_{b=1}^Q \theta_a \Theta_b r_{t-(a+bS)}. \end{aligned}$$

com funções de ligação estritamente monótonas e duas vezes diferenciáveis
 $g : (0, 1) \rightarrow \mathbb{R}$ e $h : \mathbb{R}^+ \rightarrow \mathbb{R}$, e $r_t = Y_t - \mu_t$.

iRSARMAX(p, q) \times (P, Q)_S: Estrutura

- Dinâmica parcimoniosa para λ_t : valores menores de Y_{t-1} levam a maiores valores de $\lambda_t \Rightarrow$ maior probabilidade de zeros.

$$g(\lambda_t) = \alpha_0 + \alpha_1 Y_{t-1}$$

- Casos especiais: iRSARMA(p, q) (sem regressores exógenos), iRARMAX(p, q) (não sazonal), iRARMA(p, q) (não sazonal e sem regressores exógenos).

Opções de predição (amostra interna)

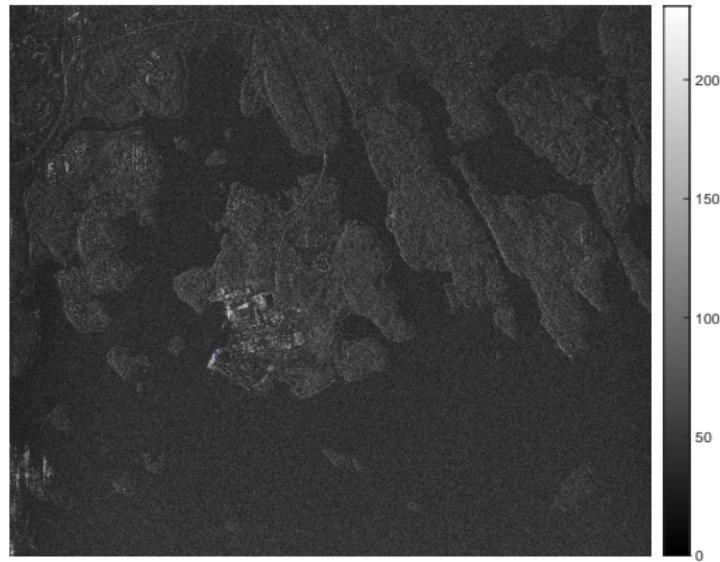
Três opções para \hat{y}_t :

$$(i) \hat{y}_t = \hat{\mu}_t \quad (\text{média da parte contínua}),$$

$$(ii) \hat{y}_t = (1 - \hat{\lambda}_t)\hat{\mu}_t \quad (\text{valor esperado da iR}),$$

$$(iii) \hat{y}_t = \begin{cases} 0, & \hat{\lambda}_t \geq 0.5, \\ \sqrt{-\frac{4\hat{\mu}_t^2}{\pi} \log\left(1 - \frac{0.5 - \hat{\lambda}_t}{1 - \hat{\lambda}_t}\right)}, & \hat{\lambda}_t < 0.5, \end{cases} \quad (\text{mediana da iR}).$$

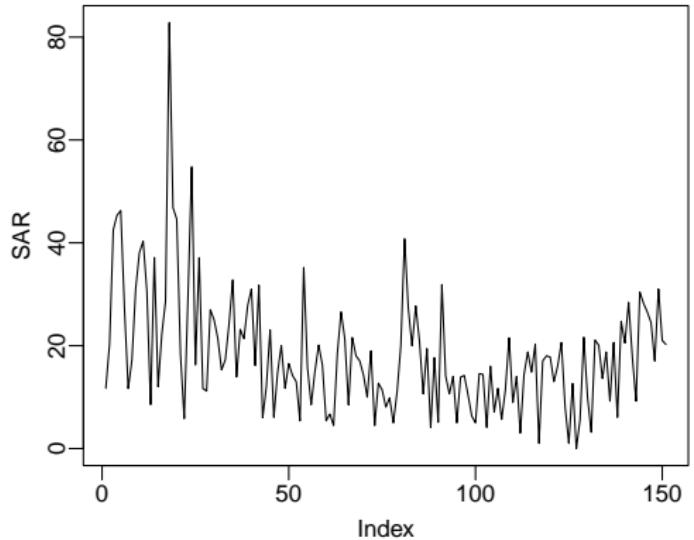
Experimento com dado SAR



Experimento com dado SAR

- **Período:** Janeiro de 2021 — coluna selecionada de uma imagem SAR.
- **Amostra e zeros:** $n = 151$ pixels; um zero (índice 127), associado a anomalias de sensor/pós-processamento.
- **Configuração do modelo:** ordens até 4, termos não sazonais e filtragem baseada em $\hat{\mu}_t$.
- **Modelo selecionado:** iRARMAX(4, 4) + $X\beta_1$ (variável indicadora de porto), com $\phi_1 = \phi_3 = \theta_1 = \theta_3 = \alpha_1 = 0$.
- **Alvo da predição:** sinal filtrado via média condicional $\hat{\mu}_t$ (parte contínua).
- **Modelos comparativos:** ARMAX (Gaussiano) e Suavização Exponencial Simples (SES).

Análise exploratória dos pixels SAR



Modelo ajustado e diagnósticos

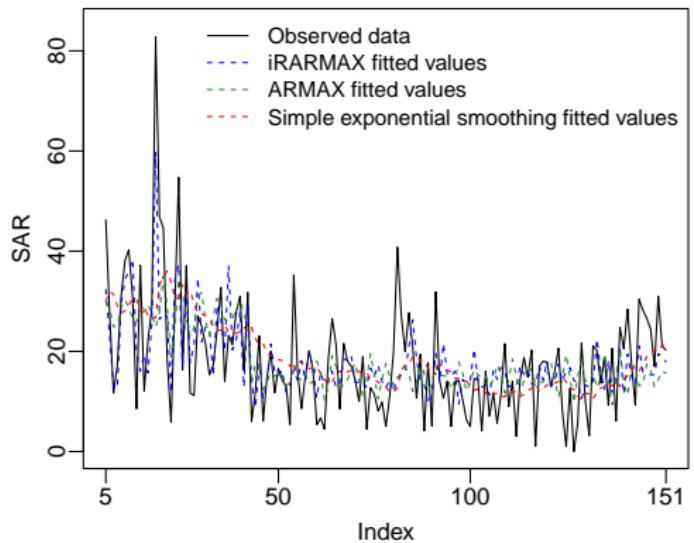
Tabela: iRARMAX(4, 4) + $X\beta_1$ ajustado.

Parâmetro	CMLE	Inferior	Superior	Wald	p-valor
β_0	3.4493	3.2660	3.6326	36.8861	< 0.0001
β_1	0.9617	0.6715	1.2518	6.4958	< 0.0001
ϕ_2	-0.0122	-0.0137	-0.0107	-16.0142	< 0.0001
ϕ_4	-0.0354	-0.0368	-0.0341	-50.5852	< 0.0001
θ_2	-0.0150	-0.0181	-0.0118	-9.4175	< 0.0001
θ_4	-0.0556	-0.0589	-0.0523	-33.3673	< 0.0001
α_0	-4.9836	-6.9503	-3.0169	-4.9666	< 0.0001

MBIC = 1088.73

Ljung–Box = 9.17 ($p=0.164$); Jarque–Bera = 1.32 ($p=0.52$); ARCH = 15.60 ($p=0.11$)

Comparação dos valores ajustados



Precisão da filtragem

Tabela: Métricas de precisão dos valores ajustados (melhor em azul).

Modelo	MAE	RMSE	MdRAE	MASE	MDA
iRARMAX	6.29	7.89	0.63	0.63	0.80
ARMAX	7.35	9.81	0.77	0.73	0.77
SES	7.67	10.41	0.79	0.76	0.73

Model ARMA 2-D Rayleigh

Motivação e Contexto

Modelagem de imagens SAR

- Pixels assumem diferentes níveis de dependência:
 - ▶ Resolução espacial;
 - ▶ Temporal.
- Um modelo ARMA 2-D é uma alternativa.

Modelo ARMA 2-D Rayleigh

Pressupostos

- Seja $Y[n, m]$, uma variável aleatória com distribuição de Rayleigh representando os pixels de uma imagem;
- Seja $y[n, m]$ a realização do sinal observado $Y[n, m]$;
- Seja

$$S[n, m] = \{[k, l] \in \mathbb{Z}^2 : 1 \leq k \leq n, 1 \leq l \leq m\} - \{[n, m]\}$$

a região fortemente causal em $[n, m]$.

Modelo ARMA 2-D Rayleigh

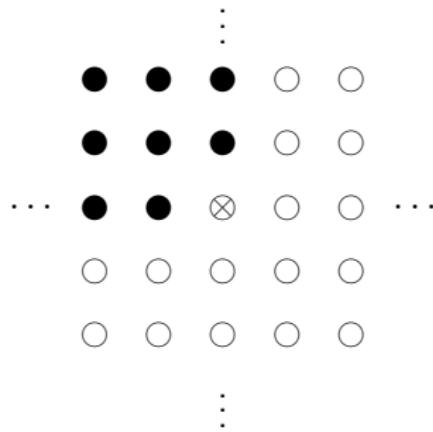


Figura: A região fortemente causal em $y[n, m]$.

Modelo ARMA 2-D Rayleigh

O modelo

$$g(\mu[n, m]) = \beta + \sum_{i=0}^p \sum_{j=0}^p \phi_{(i,j)} g(y[n-i, m-j]) + \sum_{k=0}^q \sum_{l=0}^q \theta_{(k,l)} e[n-k, m-l],$$

em que

- $\beta \in \mathbb{R}$ é um intercepto;
- p e q são as ordens do modelo;
- $\phi_{(i,j)}$, $i, j = 0, 1, \dots, p$, e $\theta_{(k,l)}$, $k, l = 0, 1, \dots, q$, são, respectivamente, os parâmetros autorregressivos e de médias móveis;
- $e[n, m] = g(y[n, m]) - g(\mu[n, m])$ é o termo de erro de médias móveis;
- $g(\mu[n, m]) = \eta[n, m]$ é o preditor linear.

Modelo ARMA 2-D Rayleigh

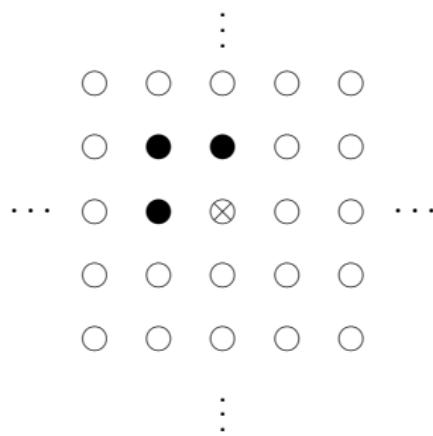


Figura: Exemplo da vizinhança utilizada em um modelo 2-D RARMA(1,1).

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias

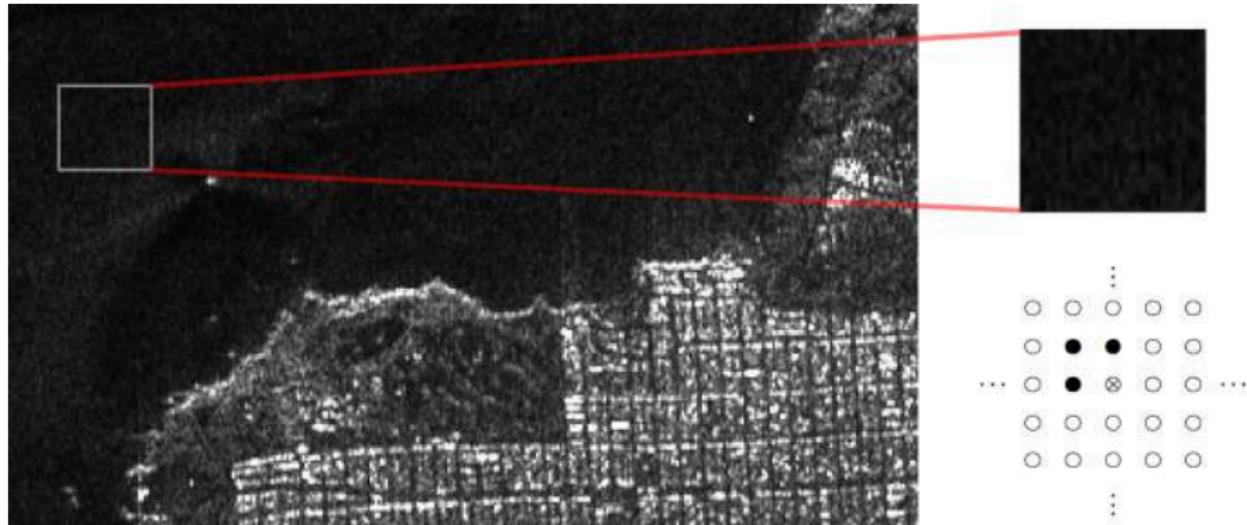


Figura: Imagem SAR de San Francisco e região de interesse.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias

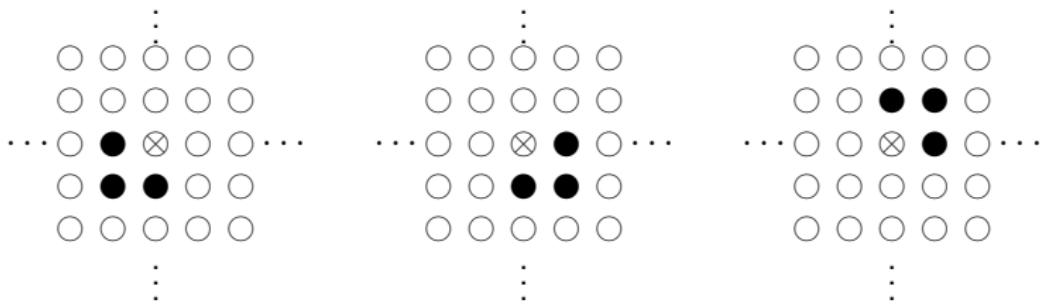


Figura: Vizinhanças utilizadas no método de detecção.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias



(a) Northwest



(b) Southwest



(c) Southeast



(d) Northeast

Figura: Imagens dos resíduos das quatro vizinhanças consideradas.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias



(a) Northwest



(b) Southwest



(c) Southeast



(d) Northeast

Figura: Imagens detectadas das vizinhanças consideradas.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias

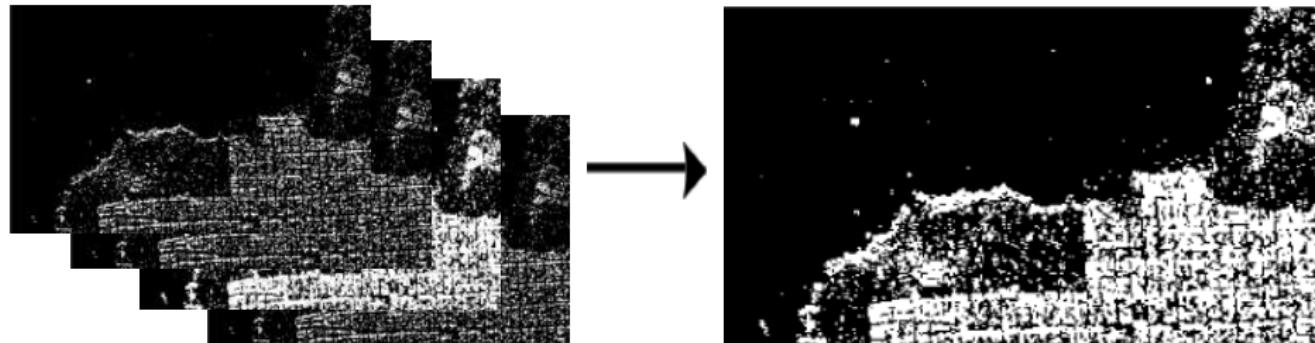


Figura: Imagens detectadas das vizinhanças consideradas.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias



(a) 2-D RARMA(1,0) model



(b) 2-D ARMA(1,0) model

Figura: Imagens detectadas com base nos modelos 2-D RARMA(1,0) e 2-D ARMA(1,0).

Modelo ARMA 2-D Rayleigh - Modelagem de Imagem SAR

Tabela: Medidas de qualidade do ajuste da imagem SAR de San Francisco com base nos modelos RARMA(1, 0) 2-D e ARMA(1, 0) 2-D.

	Modelo	
	RARMA(1, 0) 2-D	ARMA(1, 0) 2-D
MSE	0.2255	0.3191
MAPE	0.3405	0.9711

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias

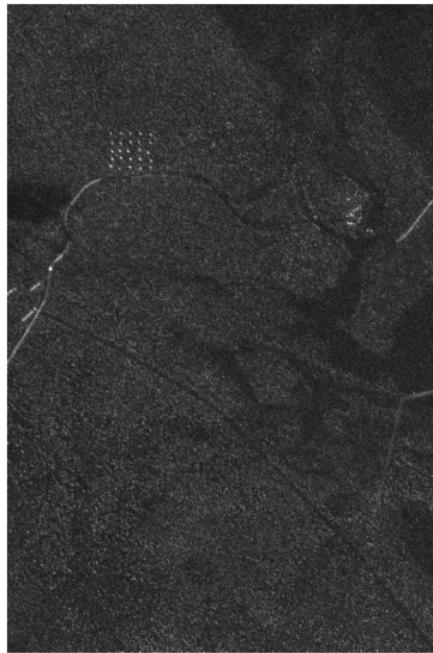
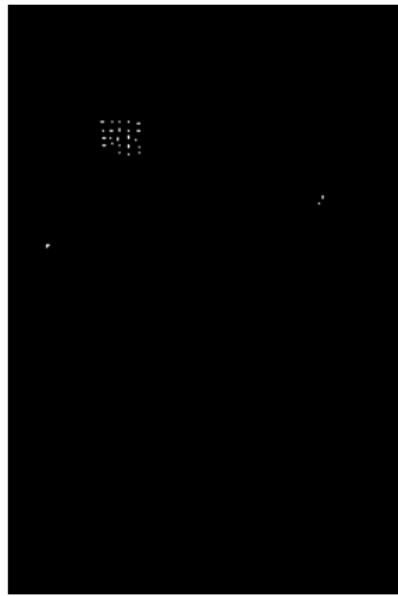
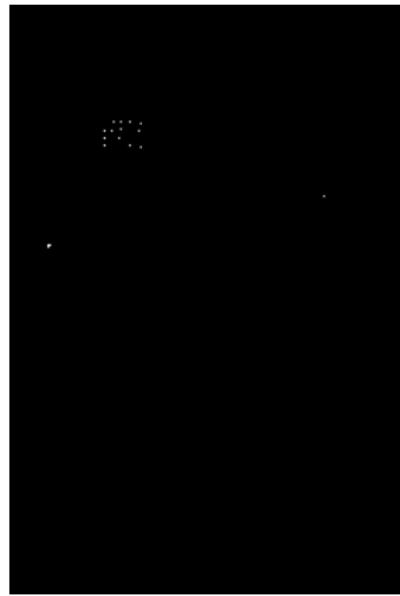


Figura: Imagem SAR original CARABAS II.

Modelo ARMA 2-D Rayleigh - Detecção de Anomalias



(a) Modelo RARMA(1, 1) 2-D



(b) Modelo ARMA(1, 1) 2-D

Figura: Imagem detectada com base nos modelos RARMA(1, 1) 2-D e ARMA(1, 1) 2-D.

Modelo ARMA 2-D Rayleigh - Modelagem de Imagem SAR

Tabela: Medidas de qualidade do ajuste da imagem SAR CARABAS II com base nos modelos RARMA(1, 1) 2-D e ARMA(1, 1) 2-D.

	Modelo	
	RARMA(1, 1) 2-D	ARMA(1, 1) 2-D
MSE	0.0562	0.1241
MAPE	0.4277	0.7499

Modelos Regressão para a Distribuição Kumaraswamy Modificada Inflacionada

Motivação

NDVI e Monitoramento da Vegetação

- O NDVI é amplamente utilizado para monitorar a saúde da vegetação.
- Seus valores são limitados em $[-1, 1]$, geralmente enviesados para valores altos.

Desafios Estatísticos

- Dados de NDVI são limitados e assimétricos, não seguem distribuição Gaussiana.
- Modelos clássicos para dados limitados (Beta, Kumaraswamy) não acomodam zeros e uns.
- Modelos inflacionados permitem combinar componentes contínuos e discretos para capturar valores de fronteira.

Motivação

Nossa Contribuição

- Proposta da **distribuição Kumaraswamy modificada inflacionada (iMK)**.
- Extensão para um modelo de regressão em nível de pixel.
- Aplicação em imagens NDVI para detecção de plantas invasoras (ex.: *black-grass* no sul da Suécia).

Distribuição iMK

Definição

- A distribuição iMK é baseada em uma mistura contínua-discreta.
- Combina a distribuição Kumaraswamy Modificada (MK) com um componente Bernoulli que adiciona massa de probabilidade em 0, 1 ou em ambos.
- Casos particulares:
 - ▶ $p = 0$: inflacionada em 0.
 - ▶ $p = 1$: inflacionada em 1.
 - ▶ $p = \lambda = 0$: reduz-se à MK padrão.

Distribuição iMK

A função distribuição de probabilidade da distribuição iMK é dada por:

$$\bar{f}(y; \vartheta, \alpha) = \begin{cases} \lambda(1 - p), & \text{if } y = 0, \\ (1 - \lambda)f(y; \mu, \alpha), & 0 < y < 1, \\ \lambda p, & \text{if } y = 1, \end{cases}$$

em que

$$f(y; \mu, \alpha) = \frac{\alpha \log(0.5) e^{\alpha - \frac{\alpha}{y}} \left(1 - e^{\alpha - \frac{\alpha}{y}}\right)^{\frac{\log(0.5)}{\mu^*} - 1}}{\mu^* y^2},$$

sendo $\mu^* = \log \left(1 - e^{\alpha - \frac{\alpha}{\mu}}\right)$ e $\alpha > 0$ o parâmetro de forma.

Modelo de Regressão

- Modelo de regressão iMK com preditores lineares:

$$g_1(\mu[n]) = \sum_{j=1}^k x_j[n] \beta_j,$$

$$g_2(\lambda[n]) = \sum_{i=1}^c z_i[n] \nu_i,$$

$$g_3(p[n]) = \sum_{b=1}^m a_b[n] \rho_b.$$

- Funções de ligação:
 - ▶ g_1 , g_2 e g_3 : logito, probito, log-log, log-log complementar, Cauchy.
- Quando a resposta assume apenas valores no intervalo aberto $(0, 1)$, surge o modelo de regressão MK, dado apenas pelo submodelo de $\mu[n]$.

Experimento com Imagem NDVI

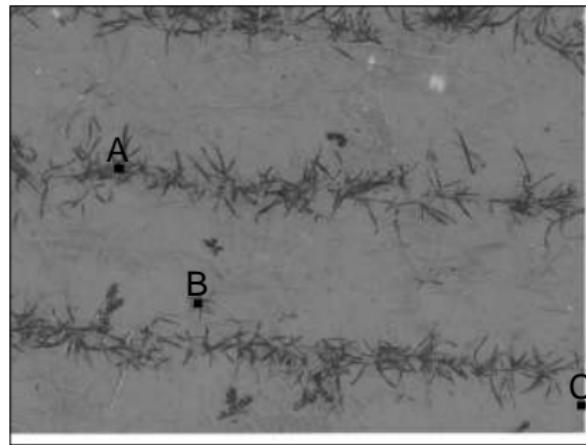


Figura: Imagem NDVI usada e áreas de interesse.

Experimento com Imagem NDVI

- O sinal de resposta foi modelado utilizando a seguinte estrutura:

$$g_1(\lambda[n]) = \gamma_1,$$

$$g_2(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n].$$

- Na equação, $y[n]$ é a magnitude vetorizada dos pixels, e $x_2[n]$ e $x_3[n]$ representam os tipos de terreno grama, peste e solo, respectivamente.
- O experimento estima λ considerando apenas o intercepto, assumindo que os valores de pixels nulos são independentes do tipo de terreno.

Experimento com Imagem NDVI

Tabela: Teste de significância para os modelos de regressão iKreg e Gaussiano aplicados aos dados dos pixels da imagem NDVI

Modelo	Parâmetro	EMV	p-valor
iMKreg	β_1	0.6812	<0.0001
	β_2	0.1722	0.0476
	β_3	-0.7170	<0.0001
iKreg	β_1	0.6978	<0.0001
	β_2	0.1664	0.0515
	β_3	-0.7454	<0.0001
Regressão Gaussiana	β_1	0.6349	<0.0001
	β_2	0.0737	0.2135
	β_3	-0.2372	0.0003

Modelo ARMA 2-D Kumaraswamy

Modelo KARMA 2-D

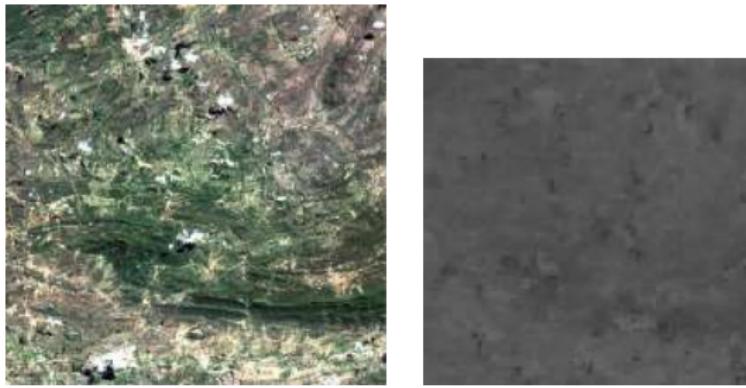
Motivação

- Em imagens de sensoriamento remoto, os pixels apresentam forte dependência espacial.
- Dados restritos ao intervalo $(0, 1)$ requerem distribuições adequadas (ex.: Kumaraswamy Modificada).
- O modelo KARMA 2-D é proposto para capturar tanto a distribuição dos dados quanto a correlação espacial.

Ideia Principal

- Extensão dos modelos ARMA bidimensionais para o caso de dados contínuos em $(0, 1)$.
- Estrutura autoregressiva e de médias móveis aplicada diretamente sobre os pixels.
- Permite analisar padrões e detectar anomalias em imagens.

Experimento com Imagem NDVI



(a) Imagem óptica

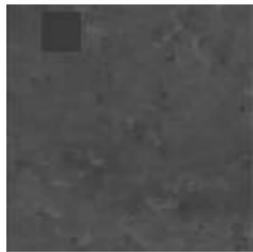
(b) Imagem NDVI

Figura: Imagem NDVI utilizada nesse experimento.

Experimento com Imagem NDVI



(a) Imagem original (b) Imagem com bloco



(c) Modelo KARMA 2-D (d) Modelo ARMA 2-D

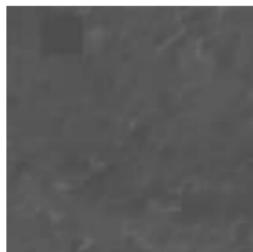
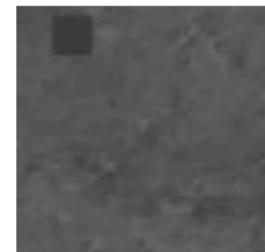


Figura: Imagem original e imagem predita pelos modelos KARMA e ARMA.

Experimento com Imagem NDVI



(a) Resíduo modelo KARMA 2-D



(b) Resíduo modelo KARMA 2-D



(c) Detecção modelo KARMA 2-D



(d) Detecção modelo ARMA 2-D

Figura: Detecção e resíduos dos modelos KARMA e ARMA.

Trabalhos em Desenvolvimento

- Teste de especificação no modelo 2-D RARMA (conferência no próximo SINAPE).
- Modelo 2-D GARMA para dados da família exponencial canônica.
- Modelo 2-D β ARMA, concorrente ao 2-D KARMA.
- Modelo 2-D CHARMA baseado na distribuição de Chen, mais flexível que a distribuição de Rayleigh.
- Modelo 2-D PBARMA baseado na distribuição Beta Prime, mais flexível que a distribuição de Rayleigh.
- Modelo 2-D WeiARMA baseado na distribuição de Weibull, da qual a distribuição de Rayleigh é um caso particular.
- Modelos dinâmicos 2-D com integração fracionária para imagens de alta resolução (memória longa).
- Generalizações para 3-D (incluindo a dimensão temporal) ou k -D, com $k > 2$, como em imagens hiperespectrais ao longo do tempo (considerando comprimento de onda e tempo como dimensões adicionais).
- Estudos de estacionaridade nesses processos (mesma estrutura em todas as dimensões?).
- Estimação robusta nesses modelos bidimensionais.
- Estudos aplicados em colaboração, abordando problemas práticos.

Nosso grupo de pesquisa

Membros principais:

- Fábio Bayer (UFSM)
- Renata Rojas Guerra (UFSM)
- Jodavid Ferreira (UFPE)
- Jairo Santana (UFSM)
- Renato Machado (ITA)
- Renato Cintra (UFPE)
- Bruna Palm (BTH)

Parceria BTH–universidades brasileiras (últimos 15 anos):

- 2 pesquisadores visitantes
- 5 pós-doutorandos na BTH e 1 na UFSM
- 6 alunos de doutorado sanduíche
- 1 aluno de doutorado pleno
- 2 alunos de doutorado (UFSM)
- alunos de mestrado e doutorado (em andamento)

Objetivos futuros

Atualmente, sou colaboradora do Programa de Pós-graduação em Modelos de Decisão e Saúde da UFPB, o que possibilita a **extensão dos modelos desenvolvidos** para aplicações em **dados da área da saúde**.

GitHub Repository

<https://github.com/palmbg>
bruna.palm@bth.se



Tack så mycket!