



INSTITUTO SUPERIOR TÉCNICO  
PROCESSAMENTO DIGITAL DE SINAIS

---

# **Deteção de Anomalias em Dados de Produção de Energia**

---

Mestrado Integrado em Engenharia Aeroespacial

Renato Loureiro, 89708  
Miguel Félix, 87083

2020/21

# 1 Estimadores de máxima verosimilhança

## 1.1 R1.a)

### Distribuição exponencial

A função de densidade probabilidade é dada por

$$p(x, \beta) = \frac{1}{\beta} e^{-\frac{x_i}{\beta}}.$$

Define-se a função logarítmica de verosimilhança

$$l(\beta) = -N \log(\beta) - \frac{\sum_{i=1}^N x_i}{\beta}$$

sendo que o parâmetro  $\beta$  é dado por

$$\hat{\beta} = \frac{\sum_{i=1}^N x_i}{N}.$$

### Distribuição de Rayleigh

A função de densidade probabilidade é dada por

$$p(x, f) = \frac{x_i}{f} e^{-\frac{x_i^2}{2f}}.$$

Define-se a função logarítmica de verosimilhança

$$l(f) = \sum_{i=1}^N \log(x_i) - N \log(f) - \frac{\sum_{i=1}^N x_i^2}{2f}$$

sendo que o parâmetro  $f$  é dado por

$$\hat{f} = \frac{\sum_{i=1}^N x_i^2}{2N}.$$

### Distribuição normal

Nesta distribuição, ao contrário das vistas anteriormente, temos dois parâmetros para estimar. A função de densidade probabilidade é dada por

$$p(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Define-se a função logarítmica de verosimilhança

$$l(\mu, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2$$

sendo que os parâmetros  $\mu$  e  $\sigma^2$  são dados por

$$\hat{\mu} = \frac{\sum_{i=1}^N x_i}{N}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (x_i - \hat{\mu})^2}{N}.$$

## 1.2 R1.b)

Depois de carregada, a imagem visualizada encontra-se na figura.

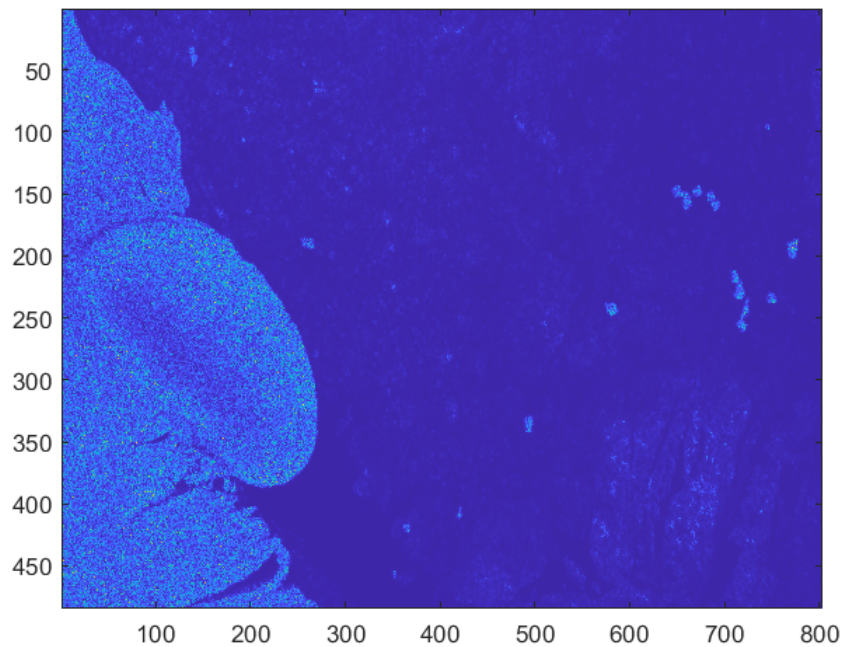
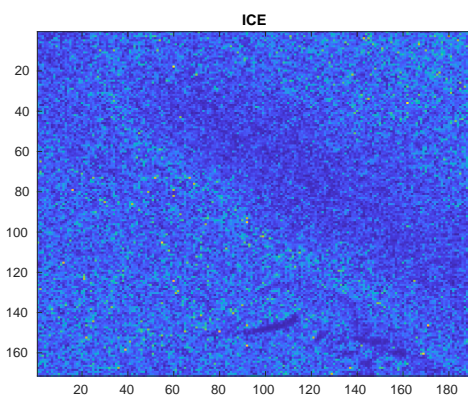
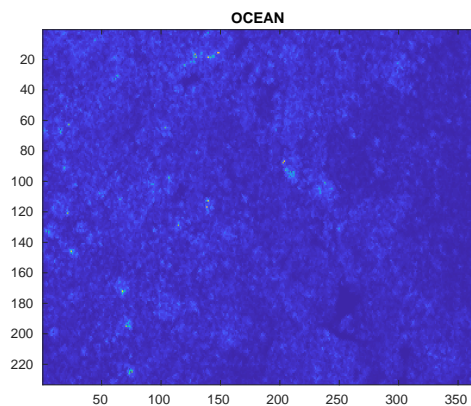


Figura 1: Imagem `sar_image.mat` visualizada com o auxílio das funções `colormap` (*parula* - 200 cores) e `imagesc`.

Tendo sido escolhidos, os seguintes pedaços para representar o gelo e o oceano.



(a) Pedaço da imagem escolhido para representar o gelo.



(b) Pedaço de imagem escolhido para representar o oceano.

Figura 2

### 1.3 R1.c)

As estimações dos parâmetros obtidas tanto para a zona de gelo como a de oceano são as demonstradas na tabela.

Distribuição	Parâmetro	Valor estimado
Rayleigh	$f$	0.1486
Exponencial	$\beta$	0.1755
Normal	$\sigma$	0.1157
	$\mu$	0.1755

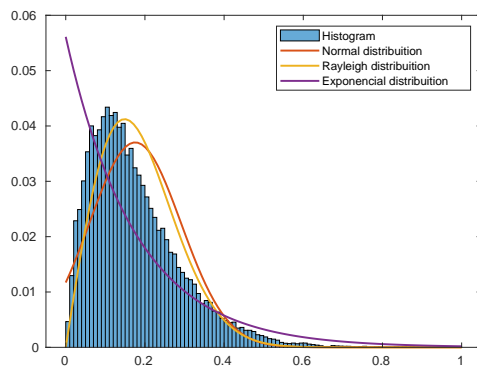
Tabela 1: Parâmetros das distribuições, referentes ao gelo.

Distribuição	Parâmetro	Valor estimado
Rayleigh	$f$	0.0406
Exponencial	$\beta$	0.0460
Normal	$\sigma$	0.0345
	$\mu$	0.0460

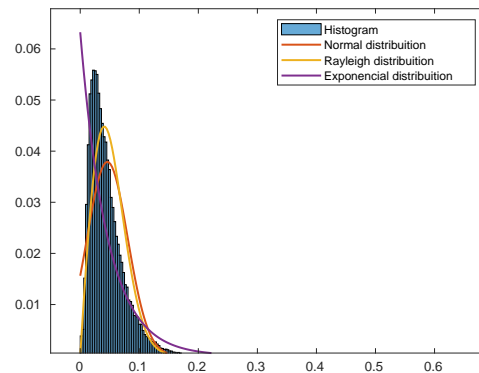
Tabela 2: Parâmetros das distribuições, referentes ao oceano.

### 1.4 R1.d)

Para averiguar qual a distribuição que melhor se adapta aos parâmetros, fizeram-se os *plots* das diferentes distribuições sobrepostas às distribuições tanto do gelo como do oceano, como demonstrado nas imagens.



(a) Referente ao gelo.



(b) Referente ao oceano.

Figura 3

Pelas imagens, conclui-se que as melhores distribuições tanto para o gelo como para o oceano são as distribuições de Rayleigh, pois são aquelas que visualmente mais se aproximam aos dados representados na forma de histograma.

## 2 Segmentação da imagem

### 2.1 R2.a)

A imagem obtida para ilustrar os resultados obtidos nesta alínea é a representada em 4.

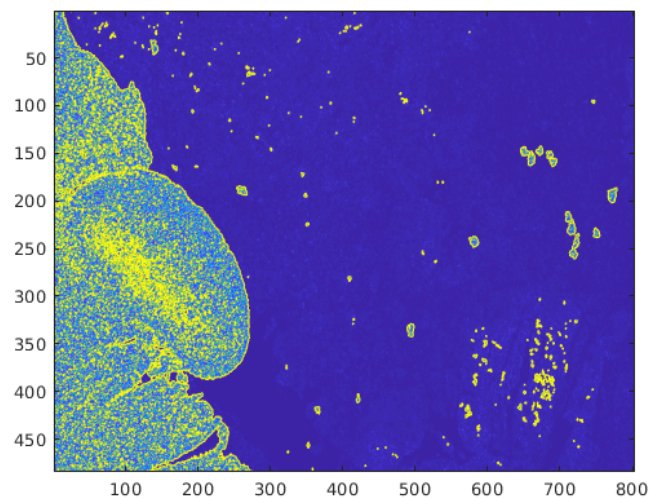


Figura 4: Segmentação da imagem usando a distribuição de Rayleigh 1x1

### 2.2 R2.b)

De igual forma, os resultados para esta alínea são apresentados na imagem 5.

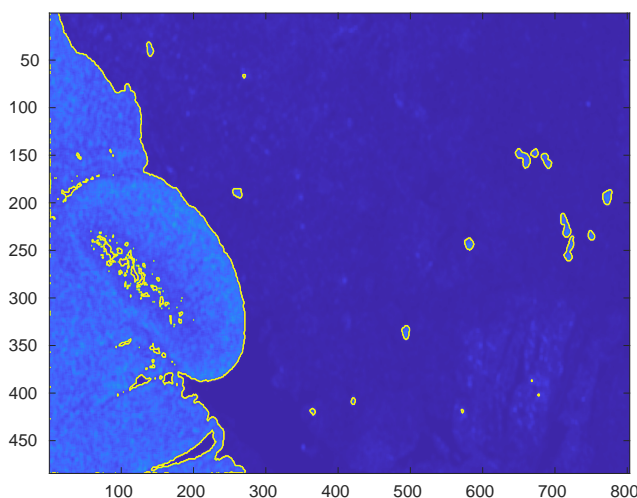


Figura 5: Segmentação da imagem usando a distribuição de Rayleigh  $5 \times 5$

## 2.3 R2.c)

Nesta secção são usados dois métodos, visual e analítico, para decidir qual das duas abordagens apresentadas anteriormente demonstra melhores resultados.

Visualmente, de uma forma global, os resultados obtidos usando a segunda abordagem (*kernel*  $5 \times 5$ ) são melhores, pois é possível analisar que o grande pedaço de gelo encostado à lateral esquerda da imagem está melhor caracterizado e corresponde com maior correcção à imagem real. Em relação ao oceano, a diferença principal é a existência de pequenas “ilhas” de gelo que estão em maior número na primeira abordagem, não sendo no entanto claro se o resultado é melhor ou pior.

Tabela 3: Tabela com os *rates* calculados para o gelo e o oceano, para *kernel*  $1 \times 1$  e  $5 \times 5$

	1x1 kernel	5x5 kernel
<b>rate</b> <sub>ICE</sub>	72.56 %	95.93 %
<b>rate</b> <sub>OCEAN</sub>	93.13 %	97.28 %

Analisando a tabela, é possível verificar que no caso do gelo, o *kernel*  $5 \times 5$  apresenta muito melhores resultados comparativamente ao de  $1 \times 1$  (um aumento de cerca de 23%). Já no caso do oceano, o mesmo pode ser dito, visto que o *rate* aumenta cerca de 4%, de 93.13% para 97.28%. Estes resultados são congruentes com a análise visual feita no parágrafo anterior, exceptuando o aumento do *rate* do oceano que, visivelmente, não é muito perceptível, mas também não surpreende.

Seria expectável que em ambos os casos houvesse melhorias quando o *kernel* de maior dimensão fosse usado, o que aconteceu em ambas as análises efectuadas.

A utilização de um *kernel* maior tem vantagens pois reduz o ruído e detecta mais facilmente os pedaços maiores de gelo e oceano, negligenciando os pequenos pontos que são visíveis na primeira abordagem. Assumindo que os pequenos pontos de gelo na seção oceânica (lateral direita) poderão ser ruído, visto que o gelo normalmente existe em grandes pedaços na natureza, a utilização de um *kernel* mais elevado revela-se vantajosa.