



Universidade Federal de Pernambuco  
Centro de Ciências Exatas e da Natureza

Pós-graduação em Estatística

Qualificação de Doutorado

## **Identifying Heterogeneity in SAR Data with New Test Statistics**

Rosa Janeth Alpala

Orientador: Dr. Alejandro C. Frery  
Co-orientador: Dr. Abraão D. C. Nascimento

Recife  
2024

# Abstract

This work presents a statistical approach to identify the underlying roughness characteristics in synthetic aperture radar (SAR) intensity data. The physical modeling of this kind of data allows the use of the Gamma distribution in the presence of fully-developed speckle, i.e., when there are infinitely many independent backscatterers per resolution cell, and none dominates the return. Such areas are often called “homogeneous” or “textureless” regions. The  $\mathcal{G}_I^0$  distribution is also a widely accepted law for heterogeneous and extremely heterogeneous regions, i.e., areas where the fully-developed speckle hypotheses do not hold. One issue involving the parametric space of  $\mathcal{G}_I^0$  is the analytical infeasibility of testing homogeneity against heterogeneity using classical tests. As solutions to this problem, we propose three test statistics to distinguish between homogeneous and inhomogeneous regions, i.e., between Gamma and  $\mathcal{G}_I^0$  distributed data, both with a known number of looks. The first test statistic uses a bootstrapped non-parametric estimator of Shannon entropy, providing an assessment in uncertain distributional assumptions. The second test uses the classical coefficient of variation (CV). The third test uses an alternative form of estimating the CV based on the ratio of the mean absolute deviation from the median to the median. We apply our test statistic to create maps of *p*-values for the homogeneity hypothesis. Finally, we show that our proposal, the entropy-based test, outperforms existing methods, such as the classical CV and its alternative variant, in identifying heterogeneity when applied to both simulated and actual data.

**Keywords:** SAR; heterogeneity; entropy; coefficient of variation; hypothesis tests

# Resumo

Este trabalho apresenta uma abordagem estatística para identificar as características de rugosidade subjacentes em dados de imagens em formação de intensidade de sensores de radar de abertura sintética (SAR – *Synthetic Aperture Radar*). A modelagem física desse tipo de dados permite o uso da distribuição Gamma na presença de speckle totalmente desenvolvido, ou seja, quando há infinitos retroespelhadores independentes por célula de resolução e nenhum domina o retorno. Essas áreas são frequentemente chamadas de regiões “homogêneas” ou “sem textura”. A distribuição  $\mathcal{G}_I^0$  também é uma lei amplamente aceita para regiões heterogêneas e extremamente heterogêneas, ou seja, áreas onde as hipóteses de speckle totalmente desenvolvidas não se aplicam. Uma problemática envolvendo o espaço paramétrico da  $\mathcal{G}_I^0$  é a inviabilidade analítica de se testar homogeneidade contra heterogeneidade a partir de testes clássicos. Como soluções para esta questão, propomos três testes de hipóteses para distinguir entre regiões homogêneas e não homogêneas, ou seja, entre dados distribuídos Gamma e  $\mathcal{G}_I^0$ , ambos com um número conhecido de looks. A primeira estatística de teste usa um estimador não paramétrico da entropia de Shannon, incorporando metodologias de bootstrap. O segundo teste usa o coeficiente de variação clássico (CV). O terceiro teste usa uma forma alternativa de estimar o CV com base na razão da média do desvio absoluto em relação à mediana. Mostramos que nossa proposta, o teste baseado em entropia, supera métodos existentes, tais como o CV clássico e sua variante alternativa, na identificação de heterogeneidade quando aplicado a dados simulados e imagens SAR. Aplicamos nossa estatística de teste para criar mapas de *p*-valores para a hipótese de homogeneidade.

**Palavras-chave:** SAR; heterogeneidade; entropia; coeficiente de variação; testes de hipóteses

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Resumo</b>	<b>iii</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Objectives	2
1.2 Manuscript organization	3
<b>2 Synthetic Aperture Radar Basics</b>	<b>4</b>
2.1 Remote Sensing	4
2.1.1 Remote Sensors	4
2.1.2 The Electromagnetic Spectrum	5
2.1.3 Scattering Mechanisms	7
2.1.4 Electromagnetic Polarization	8
2.2 SAR Image Acquisition	8
2.2.1 SAR resolution	9
2.2.2 Speckle Noise	9
2.3 Dataset	10
2.3.1 Sentinel-1 SAR data processing	10
<b>3 Methodology</b>	<b>12</b>
3.1 Statistical Modeling of Intensity SAR data	12
3.2 The Shannon Entropy	13
3.2.1 Estimation of the Shannon Entropy	14
3.2.2 Enhanced estimators with Bootstrap	16
3.3 Coefficient of variation and a robust alternative	20
3.4 Hypothesis Testing	20
3.4.1 The Proposed Test Based on Non-parametric Entropy	20
3.4.2 The Proposed Test Based on CV and a Robust Alternative	24

3.4.3 Model Selection Criterion	25
<b>4 Results</b>	<b>29</b>
4.1 Simulated Data	29
4.2 SAR Data	31
<b>5 Conclusions and future perspectives</b>	<b>36</b>
<b>Bibliography</b>	<b>37</b>

# List of Figures

2.1	Spectral regions of sensors within the electromagnetic spectrum.	6
2.2	Sentinel-1 radar.	7
2.3	Polarization in radar systems	8
2.4	Sentinel-1 radar.	9
2.5	SAR image with speckle noise	10
2.6	Selection of area of interest	11
2.7	SAR Images pre-processing in the SNAP	11
3.1	Theoretical entropies for $\Gamma_{\text{SAR}}$ and $G_I^0$ distributions.	13
3.2	$H_{G_I^0}$ converges to the $H_{\Gamma_{\text{SAR}}}$ when $\alpha \rightarrow -\infty$ , with $L = 8$ .	14
3.3	Bias and MSE of the entropy estimators for the $\Gamma_{\text{SAR}}$ , with $L = 5$ .	16
3.4	Comparing Bias and MSE: original vs. bootstrap estimators, with, $\mu = 1$ and $L = 5$ .	18
3.5	Bias and MSE of the entropy estimators for the $\Gamma_{\text{SAR}}$ , with $L = 5$ .	19
3.6	Empirical densities obtained from $S_{\tilde{H}_{\text{AO}}}(\mathbf{Z}; L)$ test under the null hypothesis.	22
3.7	Normal Q–Q plots for $n = 121$ .	23
3.8	Goodness of fit plots for evaluating the best distribution with CV data from $\Gamma_{\text{SAR}}$ (under the null hypothesis), with $n = 49$ , $L = 5$ , and $\mu = 1$ .	27
3.9	Goodness of fit plots for evaluating the best distribution with CV data from $G_I^0$ (under the alternative hypothesis), with $n = 49$ , $L = 5$ , $\mu = 1$ , and $\alpha = -3$ .	27
3.10	Goodness of fit plots for evaluating the best distribution with $CV_{\text{MnAD}}$ data from $\Gamma_{\text{SAR}}$ (under the null hypothesis), with $n = 49$ , $L = 5$ , and $\mu = 1$ .	28
3.11	Goodness of fit plots for evaluating the best distribution with $CV_{\text{MnAD}}$ data from $G_I^0$ (under the alternative hypothesis), with $n = 49$ , $L = 5$ , $\mu = 1$ , and $\alpha = -3$ .	28
4.1	Synthetic dataset.	29
4.2	Results of applying the test statistics to the simulated image.	30
4.3	Map of $p$ -values of simulated image for each test.	30
4.4	Results for a threshold of 0.05 of the $p$ -value of simulated image for each test.	31
4.5	SAR images.	31
4.6	Results of applying the test statistics to Coast of Jalisco image.	32
4.7	Map of $p$ -values of Coast of Jalisco image for each test.	32

4.8	Results for a threshold of 0.05 of the $p$ -value of Coast of Jalisco for each test.	32
4.9	Results of applying the test statistics, Illinois-Region 1.	33
4.10	Map of $p$ -values, Illinois-Region 1.	33
4.11	Results for a threshold of 0.05 of the $p$ -value, Illinois-Region 1.	33
4.12	Results of applying the test statistics, Illinois-Region 2.	34
4.13	Map of $p$ -values, Illinois-Region 2.	34
4.14	Results for a threshold of 0.05 of the $p$ -value, Illinois-Region 2.	34

# List of Tables

2.1	Different bands in microwave remote sensing	6
2.2	List of operational typical spaceborne SAR systems	7
3.1	Bias and MSE of the entropy estimators for the $\Gamma_{\text{SAR}}$ , with $L = 5$ .	17
3.2	Bias and MSE of the entropy estimators for the $\Gamma_{\text{SAR}}$ , with $L = 5$ .	19
3.3	Test accuracy and processing time for each bootstrap-improved estimator.	21
3.4	Descriptive analysis of $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$ , with $L \in \{3, 5, 8, 11\}$ and $\mu = 1$ .	22
3.5	Size and Power of the $S_{\tilde{H}_{AO}}(\mathbf{Z})$ test statistic.	24
3.6	AIC and BIC values for evaluating the best distribution with CV data from $\Gamma_{\text{SAR}}$ .	25
3.7	AIC and BIC values for evaluating the best distribution with CV data from $G_I^0$ .	26
3.8	AIC and BIC values for evaluating the best distribution with CV <sub>MnAD</sub> data from $\Gamma_{\text{SAR}}$ .	26
3.9	AIC and BIC values for evaluating the best distribution with CV <sub>MnAD</sub> data from $G_I^0$ .	26

# Introduction

The technology of Synthetic Aperture Radar (SAR) imaging operates on the comprehensive aperture principle to produce high-resolution SAR images. In contrast to conventional optical remote sensing images, SAR images remain unaffected by external factors, making them suitable for identifying ground targets in diverse weather conditions and expansive areas (Mu et al., 2019). Consequently, SAR imagery has become essential for environmental monitoring (Amitrano et al., 2021), crop mapping (Dingle Robertson et al., 2020), ship target detection (Wang et al., 2019), and climatology (Zhao et al., 2023). However, the effective use of SAR data depends on a thorough understanding of its statistical properties because it is corrupted by speckle. This noise-like interference effect is inherent in SAR data due to the coherent nature of the imaging process (Argenti et al., 2013).

Speckle in intensity format is non-Gaussian. Thus, SAR data require reliable statistical models for accurate processing. The  $\mathcal{G}^0$  distribution, which is suitable for SAR data, includes the Gamma law as the limiting case for fully-developed speckle (De A. Ferreira et al., 2020) and provides flexibility with fewer parameters for analysis.

When deciding which model is the best, practitioners face a problem. On the one hand, if they opt for the Gamma law when the data come from the  $\mathcal{G}^0$  distribution, they lose all the information about the number of scatterers, which is revealed by one of the parameters of the latter model (Yue et al., 2021). On the other hand, if they apply the  $\mathcal{G}^0$  distribution under fully developed speckle, maximum likelihood estimation is tricky: bias increases making estimation unreliable (Vasconcellos et al., 2005), and the likelihood is flat, so numerical optimization may not converge (Frery et al., 2004).

Our work aims to improve the identification of potential roughness features in SAR intensity data. Physical modeling of SAR data allows the use of the Gamma distribution in the presence of fully-developed speckle, where an infinite number of independent backscatterers per resolution unit is assumed, commonly referred to as homogeneous regions.

Entropy is a fundamental concept in information theory with far-reaching applications in pattern recognition (Avval et al., 2021), statistical physics (Pressé et al., 2013), image processing (Mohammad-Djafari, 2015), edge detection (Nascimento et al., 2014) and SAR image analysis (Nascimento et al., 2019). Shannon (1948) introduced this concept for a random variable

to measure information and uncertainty. Shannon entropy is a crucial descriptive parameter in statistics, especially for evaluating data dispersion and performing tests for normality, exponentiality and uniformity (Wieczorkowski et al., 1999; Zamanzade et al., 2012).

Entropy estimation is challenging, especially when the model is unknown. In these cases, non-parametric methods are used. Spacing methods have been discussed as a non-parametric approach in Refs. (Noughabi, 2010; Subhash et al., 2021). This strategy is flexible and robust because it does not enforce a model or parametric constraints. In this context, we introduce a novel approach: a bootstrap-improved non-parametric estimator for Shannon entropy.

The coefficient of variation (CV), introduced in 1896 by Pearson et al. (1896), is a relative dispersion measure widely used in various fields of applied statistics, including sampling (Hendricks, 1936), biostatistics (Tian, 2005), medical and biological research (Subrahmanyam et al., 2003), climatology (Chankham et al., 2024) and other fields. It facilitates the comparison of variability between different populations and is particularly valuable for relating variables with different units. The variable with the highest CV value has the largest relative dispersion around the mean value (Banik et al., 2011). The coefficient of variation is the primary measure of heterogeneity in SAR data (Ulaby et al., 1986; Touzi et al., 1988). In our work, we employ both the classical CV and a robust alternative, which is based on the ratio of the mean absolute deviation from the median (MnAD) to the median. These measures offer a more detailed view of data variability.

The expected value, a critical component in both CV and the Shannon entropy, influences their interpretation. In CV calculations, the expected value serves as a reference point for assessing data dispersion relative to its average value, while in the Shannon entropy, the expected value also plays an important role, particularly within parametric distributions such as the Gamma or the  $G_I^0$  distribution. In these cases, the mean appears in the analytical entropy formula, indicating that the entropy of the distribution is influenced by the mean value of the data.

When applying these statistical concepts to SAR image analysis, we aim to discern between homogeneous and non-homogeneous regions, contributing to the improved interpretation and understanding of SAR data. In our study, we devise test statistics based on these three estimators: the classical coefficient of variation, a robust version, and the Shannon entropy estimator. We apply these test statistics to generate maps of evidence of homogeneity that reveal different types of targets in the SAR data, assuming the number of looks is known. Likewise, we show that our proposed method is superior to existing approaches with simulated data and SAR images.

## 1.1 Objectives

The aim of this work is the identification of roughness features in SAR intensity data by developing novel test statistics for distinguishing between homogeneous and heterogeneous domains. To achieve this goal, we propose the following procedure:

- Select the optimal non-parametric estimator of entropy, refined through bootstrap techniques, aiming to reduce bias and mean square error.
- Propose three test statistics: the first based on the selected non-parametric estimator of entropy; the second utilizing the classical coefficient of variation; and the third using a

robust approach of CV variant.

- Conduct computational experiments to evaluate the performance of the proposed test statistics across various simulated data scenarios.
- Apply the test statistics to SAR data to assess their effectiveness in identifying roughness features.

## Submitted Articles

The results obtained in this work were submitted for publication as follows:

- MIGARS 2024 proceedings in the IEEE Xplore Digital Library®, titled "Identifying Departures from the Fully Developed Speckle Hypothesis in Intensity SAR Data with Non-Parametric Estimation of the Entropy".
- Remote Sensing journal from MDPI, titled "Identifying Heterogeneity in SAR Data with New Test Statistics".

The articles were written in Rmarkdown and are fully reproducible. The code and data are accessible at [Repository Link](#)

## 1.2 Manuscript organization

This document is organized as follows: Chapter 2 provides a background on remote sensing and SAR images. Chapter 3 discusses methodological aspects, including statistical modeling, entropy estimation for intensity SAR data, and hypothesis testing. Chapter 4 presents the main results obtained in our study with both simulated and actual data. Finally, In Chapter 5, the conclusions and future work are presented.

# Synthetic Aperture Radar Basics

This chapter presents some basics of remote sensing that we will use in the rest of the document and focuses on the theory of SAR.

## 2.1 Remote Sensing

Remote sensing is the practice of deriving information about the Earth's land using images acquired from elevated platforms like airplanes or satellites in space, using electromagnetic radiation across various regions of the electromagnetic spectrum (Campbell et al., 2011).

The word “radar” was originally an acronym, RADAR, for “radio detection and ranging”. Today, the technology is so common that the word has become a standard English noun. The history of radar extends to the early days of modern electromagnetic theory. In 1886, Hertz demonstrated reflection of radio waves, and in 1900 Tesla described a concept for electromagnetic detection and velocity measurement in an interview. Early radar development was driven by military necessity, and the military is still a major user and developer of radar technology. Radar is also an essential sensor for emerging autonomous driving systems. Finally, spaceborne and airborne radar is an important tool in mapping earth topology and environmental characteristics such as water and ice conditions, forestry conditions, land usage, and pollution (Richards, 2005).

In 1951 Carl Wiley developed synthetic aperture radar while working at the Goodyear Aircraft Company. Using a technique known as Doppler beam sharpening, the signals from a series of locations are summed to produce a finer azimuth resolution than the antenna beamwidth can achieve (Wiley, 1985). Since then, the technology has found a multitude of applications in the scientific, military, and civilian communities alike. Due to the numerous uses, SAR technology has been a topic of constant research and development.

### 2.1.1 Remote Sensors

Remote sensors collect data remotely by measuring electromagnetic (EM) radiation within specific spectral ranges, often called bands. When they are deployed on satellites or mounted on aircraft, they are called spaceborne and airborne remote sensors, respectively. Spaceborne

remote sensors orbit the Earth at heights of 500 to 800 km (Everaerts, 2005), providing data in a wide range of ground resolutions, and at various polarizations and radar bands.

There are two types of remote sensors acquiring information in fundamentally different ways. Optical sensors are passive imaging devices, which means they take advantage of existing sources of illumination and the natural reflective properties of objects to form images. In the case of optical Earth observation imagery, the sun is used as a global illumination source. Thus, optical image sensors tend to operate on a sun-synchronous orbit to ensure the scene is well illuminated during acquisition. On the other hand, SAR uses the principle of active sensing, whereby the sensor acts as both an illumination source as well as the acquisition unit. A scene is thus imaged by alternating between emitting bursts of EM-radiation and measuring the strength and time delay of the reflected signal. The proportion of the signal which is reflected towards the sensor by an object is known as *backscatter*.

### 2.1.2 The Electromagnetic Spectrum

Given that objects interact (reflect and absorb) different parts of the EM-spectrum in a unique manner, several realizations of optical and SAR sensors exist, each designed for a specific purpose and to take advantage of the properties observable within a particular part of the EM-spectrum. Optical sensors detect very high-frequency radiation within the visible to thermal infrared section of the EM-spectrum. In contrast, SAR sensors operate at much lower frequencies within the microwave section of the spectrum. It is this use of lower frequencies which allow SAR sensors to be mostly independent of atmospheric conditions. The same, however, is not true for optical sensors which require clear atmospheric conditions to observe ground level reflections within the visible light and near-infrared spectrum.

In general, optical sensors can be classified by the number of spectral bands they image into hyperspectral, multispectral, and panchromatic sensors. Hyperspectral sensors image the full optical subsection of the EM spectrum into hundreds of narrow spectral bands, while panchromatic sensors image the visible light spectrum using a single, wide spectral band (Sara et al., 2021). Similarly, SAR sensors can also be classified based on their operational bandwidth into C, L and X-band, with L-band being the lowest frequency and thus the highest level of vegetation and soil penetration. However, due to the relationship between wave-length and spatial resolution present in SAR imagery (Cumming et al., 2005), X-band imagery exhibits the highest spatial resolution and thus provides the detailed information about surface structure. In optical imagery, a similar trade-off exists, except it is between spatial resolution and spectral resolution. Due to data storage and throughput constraints hyperspectral and multispectral images tend to have a lower resolution than panchromatic imagery.

The operational frequency and bandwidth of each of these sensors within the EM-spectrum is further described by Figure 2.1.

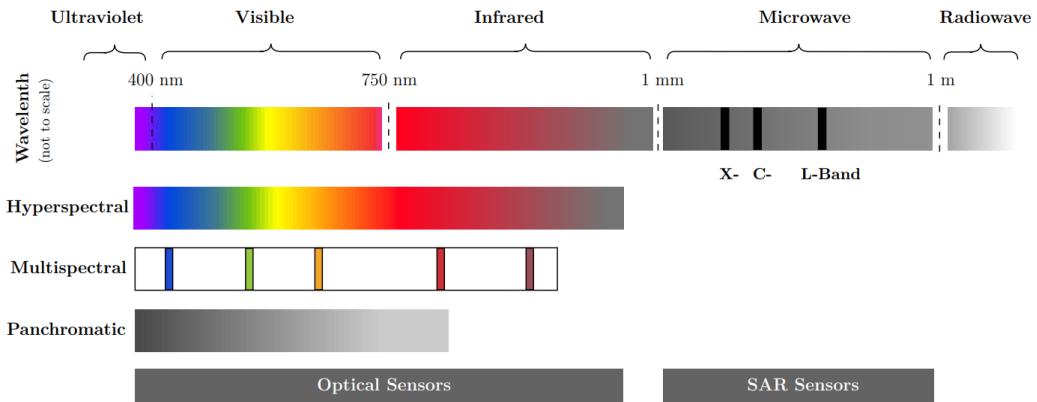


Figure 2.1: Spectral regions of optical sensors and radar sensors within the electromagnetic spectrum, (Zhang, 2022).

SAR sensors provide radar images of the Earth's surface, which are chosen in this work. Table 2.1 contains the band with the associated frequency and wavelength. The wavelength range is very important, since it determines the interaction between radar signals and surfaces, as well as the penetration depth of the microwave signals.

Table 2.1: Different bands in microwave remote sensing, (Moreira et al., 2013).

<b>Band</b>	<b>Frequency [GHz]</b>	<b>Wavelength [cm]</b>
Ka	25.0–40.0	0.75–1.2
Ku	12.0–17.6	1.7–2.5
X	7.5–12.0	2.5–4.0
C	3.7–7.5	4.0–8.0
S	2.0–3.7	8.0–15
L	1.0–2.0	15–30
P	0.3–1.0	30–100

Table 2.2 gives some examples of commonly used spaceborne SAR sensors, currently in operation. These include Radarsat-2 from the Canadian Space Agency (CSA), TerraSAR-X from the German Aerospace Center (DLR), ALOS PALSAR-2 from the Japan Aerospace Exploration Agency (JAXA), and Sentinel-1A/1B from the European Space Agency (ESA) (see Figure 2.2), as well as Gaofen-3 from the China National Space Administration (CNSA).

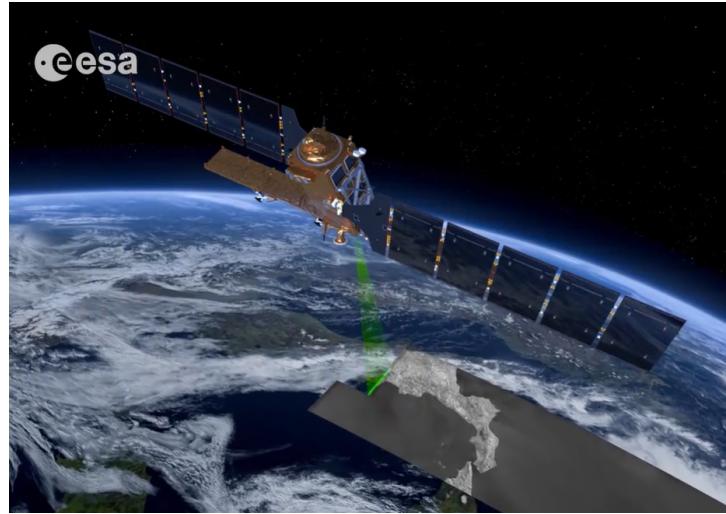


Figure 2.2: Sentinel-1 radar, (source ESA).

Table 2.2: List of operational typical spaceborne SAR systems.

Mission/SAR	Agency	Launch	Band	Resolution [m]	Coverage [km]	Revisit [days]
Radarsat-2	CSA	2007	C	9–100	18–500	24
TerraSAR-X	DLR	2007	X	0.25–40	10–150	11
ALOS PALSAR-2	JAXA	2014	L	1–100	25–490	14
Sentinel 1A/1B	ESA	2014/2016	C	5–100	80–400	12
Gaofen-3	CNSA	2016	C	1–500	10–650	29

### 2.1.3 Scattering Mechanisms

A SAR sensor measures the electromagnetic energy that is backscattered from the targets. There are lots of factors that can affect the radar backscatter, and they can be divided into two categories.

The first category relates to the SAR parameters, such as the frequency, polarization, and incident angle. If the frequency and polarization parameters are fixed, the increasing incident angle causes the decreasing backscatter intensity from a homogeneous surface. Therefore, the intensity decreases gradually on SAR images from near range to far range. This effect must be taken into consideration during SAR image interpretation.

The second category relates to the surface parameters, such as the surface roughness, the surface geometry, and the dielectric constant of the surface. There are three basic groups of scattering mechanisms that can contribute to the returned signal: surface scattering, volume scattering, and hard target scattering.

### 2.1.4 Electromagnetic Polarization

An electromagnetic wave consists of a magnetic and an electric field. These two fields are perpendicular to each other and also to the direction of wave propagation.

Apart from frequency, amplitude, and phase, an electromagnetic wave also contains polarization information. Polarization is defined as the orientation of the oscillating electric field, which can be described in terms of two orthogonal basis vectors. Electromagnetic waves are generally elliptically polarized, with linear or circular polarization as special cases.

Most of the SAR sensors use linear polarization on both the transmitter and the receiver. There are four linear polarization configurations in total:

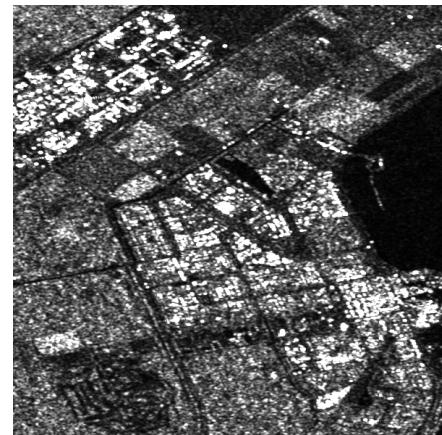
- HH (co-polarization): horizontal transmission and horizontal reception.
- HV (cross-polarization): horizontal transmission and vertical reception.
- VH (cross-polarization): vertical transmission and horizontal reception.
- VV (co-polarization): vertical transmission and vertical reception.

Historical SAR satellites carried single-polarized sensors, which support only one linear polarization. More recent sensors provide either dual-polarization or quad-polarization capabilities. For quad-polarization SAR systems, they can transmit H- and V-polarized waveforms and receive both H and V simultaneously.

Figure 2.3 (a)–(b) shows VH and VV polarization from Sentinel-1 radar images in C-Band of a region of Flevoland.



(a) Cross-polarization (VH)



(b) Co-polarization (VV)

Figure 2.3: Polarization in radar systems

## 2.2 SAR Image Acquisition

SAR sensors utilize a single antenna to emit EM-signals and measure the corresponding magnitude, range and Doppler-shift of these signals in the backscatter. As only a single detector element exists, SAR is based around the concept of synthesizing an aperture in the azimuth direction in order to form a 2-dimensional image (Orth, 2018). The angle between the SAR

antenna and an object on the ground is known as the incidence angle,  $\theta$ , and it plays a large role in the geometric distortions which are present in SAR imagery. Due to this imaging concept, the SAR image is formed along the slant range, where each pixel represents the accumulation of backscatter from points which are the same distance away from the sensor (Cumming et al., 2005). The acquisition geometry of SAR sensors is depicted in Figure 2.4.

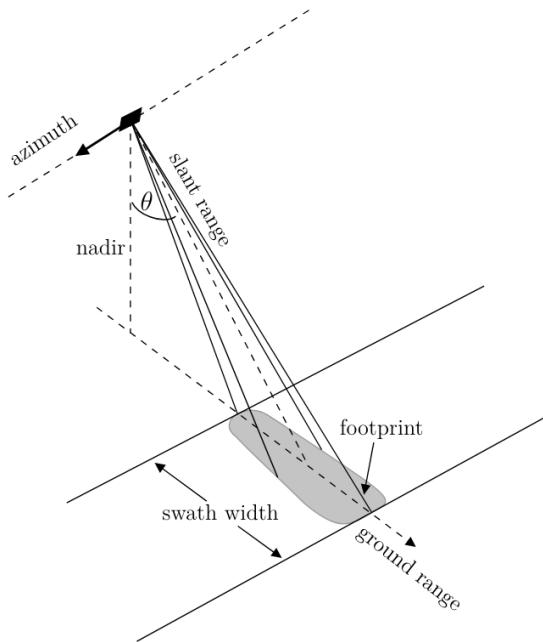


Figure 2.4: SAR image acquisition geometry, (Hughes, 2020).

### 2.2.1 SAR resolution

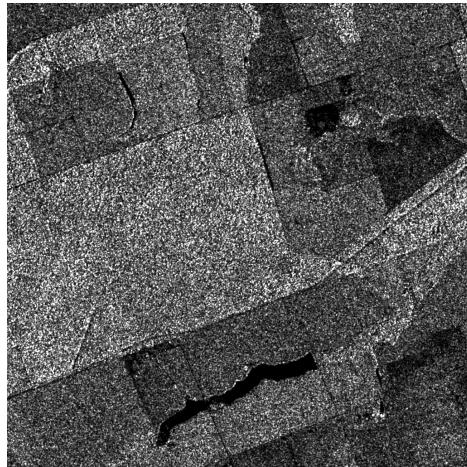
The resolution of the SAR image is one of the most important characteristics of the SAR system. SAR data pixels are characterized by spatial and temporal resolutions. Spatial resolution is defined as the ability of the SAR system to distinguish between two closely targets. In SAR systems, there are two resolutions, azimuth resolution, and range resolution. Another resolution that should be taken into account when applying a multi-temporal time series analysis of the SAR data is the temporal resolution. It is determined by the revisit time of the SAR sensor at the same imaged area.

### 2.2.2 Speckle Noise

The interference of waves reflected during the acquisition process of SAR images give rise to a multiplicative and non-Gaussian noise that characterizes them and is known as speckle noise (Oliver et al., 2004). Since many elemental scatterers are located in one resolution cell, the final scattering response from the resolution cell is the coherent sum of thousands of individual scattering events. The mutual interference results in a certain fluctuation in the amplitude and phase of the synthesized EM wave vectors, which make the “salt-and-pepper” noise appear.

The statistical distribution of the speckle is well known under certain conditions. Three main cases can be considered: homogeneous, heterogeneous, and extremely heterogeneous areas. Homogeneous areas (such as fields, pasture, water bodies, etc) are characterized by the lack of dominant scatterers, and the surface can be considered stationary. This is the case of the *fully developed hypothesis for the speckle* (Frery et al., 1997).

In Figure 2.5, SAR images depict speckle, a characteristic of SAR data due to its coherent imaging process. The images distinguish between fully developed speckle in homogeneous regions (Figure 2.5 (a)) and heterogeneous clutter in regions such as urban areas (Figure 2.5 (b)).



(a) Fully developed speckle (Homogeneous areas)



(b) Heterogeneous clutter

Figure 2.5: SAR image with speckle noise

## 2.3 Dataset

### 2.3.1 Sentinel-1 SAR data processing

The Sentinel-1 images were obtained from the Alaska Satellite Facility (ASF) at <https://search.asf.alaska.edu/#/> and by the European Space Agency via Copernicus Data Space Ecosystem at <https://dataspace.copernicus.eu/> (see Figure 2.6).

The SAR images downloaded were pre-processed using the Sentinel Application Platform (SNAP) Toolbox developed by ESA (see Figure 2.7).

The following steps were applied to the SAR data during pre-processing:

1. Importing the SAR image into SNAP.
2. Creating a subset to focus on the area of interest.
3. Performing Geometric Correction, which includes Range Doppler Terrain Correction.
4. Exporting the processed data in ENVI format for further analysis.

## 2.3 DATASET

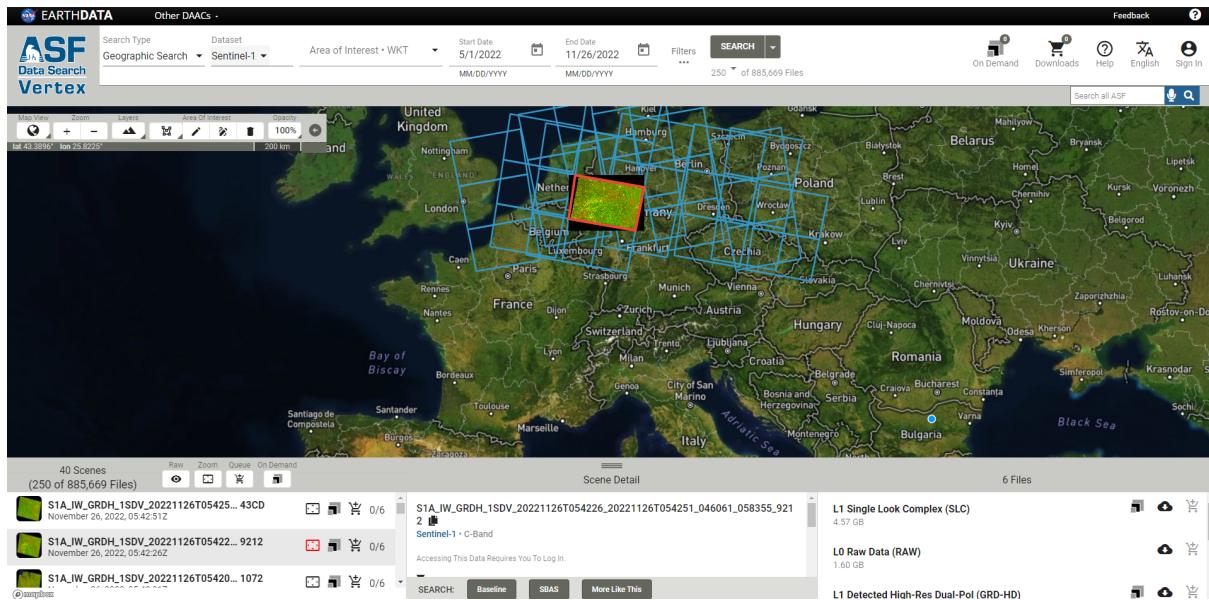


Figure 2.6: Selection of area of interest.

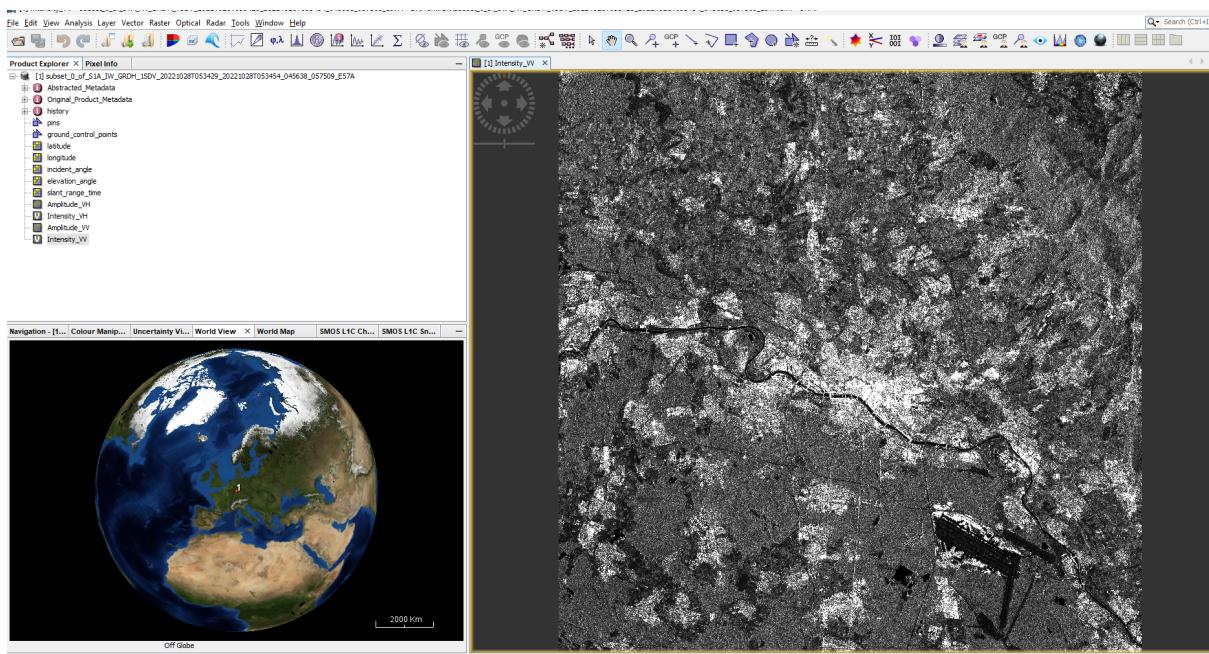


Figure 2.7: SAR Images pre-processing in SNAP.

# Methodology

## 3.1 Statistical Modeling of Intensity SAR data

The primary models for intensity SAR data include the Gamma and  $\mathcal{G}_I^0$  distributions (Frery et al., 1997). The first is suitable for fully developed speckle and is a limiting case of the second model. This is interesting due to its versatility in accurately representing regions with different roughness properties (Cassetti et al., 2022). We denote  $Z \sim \Gamma_{\text{SAR}}(L, \mu)$  and  $Z \sim \mathcal{G}_I^0(\alpha, \gamma, L)$  to indicate that  $Z$  follows the distributions characterized by the respective probability density functions (pdfs):

$$f_Z(z; L, \mu \mid \Gamma_{\text{SAR}}) = \frac{L^L}{\Gamma(L)\mu^L} z^{L-1} \exp\{-Lz/\mu\} \mathbb{1}_{\mathbb{R}_+}(z) \quad (3.1)$$

and

$$f_Z(z; \alpha, \gamma, L \mid \mathcal{G}_I^0) = \frac{L^L \Gamma(L - \alpha)}{\gamma^\alpha \Gamma(-\alpha) \Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + Lz)^{L-\alpha}} \mathbb{1}_{\mathbb{R}_+}(z), \quad (3.2)$$

where  $\mu > 0$  is the mean,  $\gamma > 0$  is the scale,  $\alpha < 0$  measures the roughness,  $L \geq 1$  is the number of looks,  $\Gamma(\cdot)$  is the gamma function, and  $\mathbb{1}_A(z)$  is the indicator function of the set  $A$ .

The  $r$ th order moments of the  $\mathcal{G}_I^0$  model are

$$E(Z^r \mid \mathcal{G}_I^0) = \left(\frac{\gamma}{L}\right)^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)} \cdot \frac{\Gamma(L + r)}{\Gamma(L)}, \quad (3.3)$$

provided  $\alpha < -r$ , and infinite otherwise. Therefore, assuming  $\alpha < -1$ , its expected value is

$$\mu = \left(\frac{\gamma}{L}\right) \frac{\Gamma(-\alpha - 1)}{\Gamma(-\alpha)} \cdot \frac{\Gamma(L + 1)}{\gamma(L)} = -\frac{\gamma}{\alpha + 1}. \quad (3.4)$$

Although the  $\mathcal{G}_I^0$  distribution is defined by the parameters  $\alpha$  and  $\gamma$ , in the SAR literature (Nascimento et al., 2010) the texture  $\alpha$  and the mean  $\mu$  are usually used. Reparametrizing (3.2) with  $\mu$ , and denoting this model as  $G_I^0$  we obtain:

$$f_Z(z; \mu, \alpha, L \mid G_I^0) = \frac{L^L \Gamma(L - \alpha)}{[-\mu(\alpha + 1)]^\alpha \Gamma(-\alpha) \Gamma(L)} \frac{z^{L-1}}{[-\mu(\alpha + 1) + Lz]^{L-\alpha}}. \quad (3.5)$$

### 3.2 The Shannon Entropy

The parametric representation of Shannon entropy for a system described by a continuous random variable is:

$$H(Z) = - \int_{-\infty}^{\infty} f(z) \ln f(z) dz, \quad (3.6)$$

here,  $f(\cdot)$  is the pdf that characterizes the distribution of the real-valued random variable  $Z$ .

Using (3.6), we obtain the Shannon entropy of  $\Gamma_{\text{SAR}}$  in (3.1) and  $G_I^0$  in (3.5):

$$H_{\Gamma_{\text{SAR}}}(L, \mu) = L - \ln L + \ln \Gamma(L) + (1 - L)\psi^{(0)}(L) + \ln \mu, \quad (3.7)$$

$$\begin{aligned} H_{G_I^0}(\mu, \alpha, L) &= \underbrace{L - \ln L + \ln \Gamma(L) + (1 - L)\psi^{(0)}(L) + \ln \mu - \ln \Gamma(L - \alpha)}_{H_{\Gamma_{\text{SAR}}}} \\ &\quad + (L - \alpha)\psi^{(0)}(L - \alpha) - (1 - \alpha)\psi^{(0)}(-\alpha) + \ln(-1 - \alpha) + \ln \Gamma(-\alpha) - L, \end{aligned} \quad (3.8)$$

where  $\psi^{(0)}(\cdot)$  is the digamma function.

Figure 3.1 (a) shows the theoretical entropy  $H_{\Gamma_{\text{SAR}}}(L, \mu)$  as a function of  $\mu$  for different values of  $L$ , while Figure 3.1 (b) shows the theoretical entropy  $H_{G_I^0}(\mu, \alpha, L)$  as a function of both  $\mu$  and  $\alpha$ , likewise considering varying  $L$ . We see that the entropy of a random variable

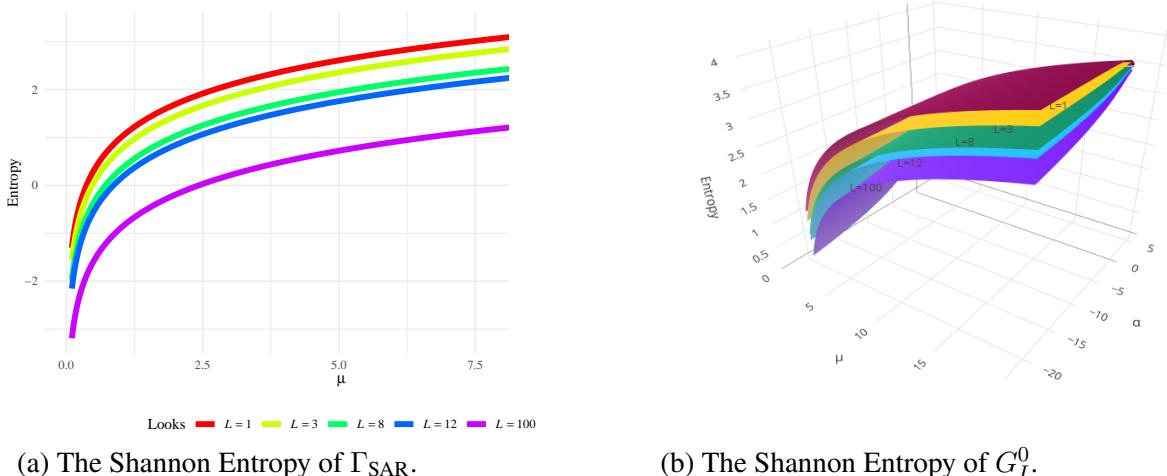


Figure 3.1: Theoretical entropies for  $\Gamma_{\text{SAR}}$  and  $G_I^0$  distributions.

following the  $\Gamma_{\text{SAR}}$  model depends on the logarithm of the mean  $\mu$ . From the 3D plot, we observe how variations in  $\mu$ ,  $\alpha$ , and  $L$  collectively affect the entropy of the  $G_I^0$  distribution, allowing the identification of regions of high or low entropy.

Additionally, the expression (3.8) can be written as:  $H_{G_I^0}(\mu, \alpha, L) = H_{\Gamma_{\text{SAR}}} + \phi(\alpha, L)$ . Thus, for  $L \geq 1$  known, and as  $\alpha \rightarrow -\infty$  we have that  $H_{G_I^0} \approx H_{\Gamma_{\text{SAR}}}$ . Figure 3.2, shows the behavior

of the entropy of  $G_I^0$  as a function of  $\mu$  when  $\alpha \in \{-\infty, -20, -8, -3\}$  and  $L = 8$ , where the approximation to the entropy of  $\Gamma_{\text{SAR}}$  can be observed when  $\alpha$  assumes large negative values. In an interpretative sense, we can conclude that the more heterogeneous (high values of  $\alpha$ ) the SAR region is, the higher the value of entropy (or degree of disorder). It is worth noting that this conclusion confirms what is written in the SAR literature.

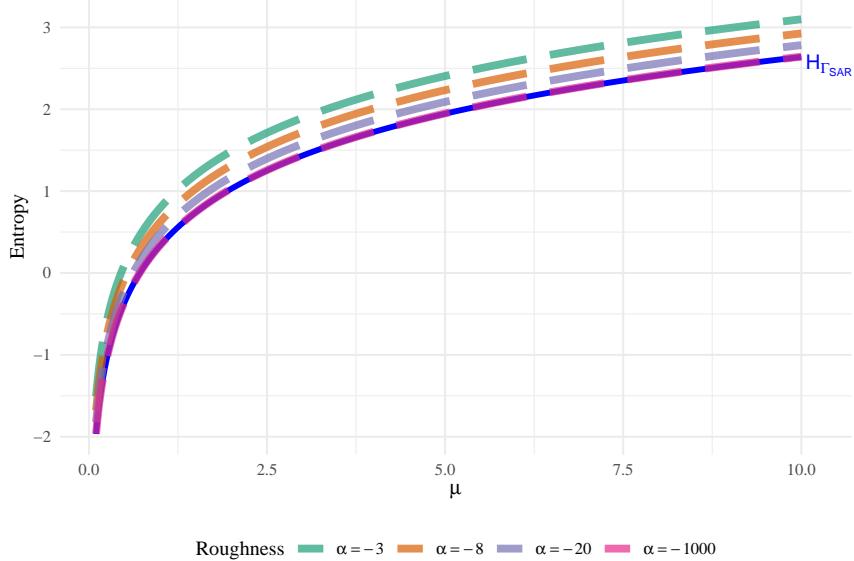


Figure 3.2:  $H_{G_I^0}$  converges to the  $H_{\Gamma_{\text{SAR}}}$  when  $\alpha \rightarrow -\infty$ , with  $L = 8$ .

### 3.2.1 Estimation of the Shannon Entropy

The problem of the non-parametric estimation of  $H(Z)$  has been studied by many authors, including Vasicek (1976), Correa (1995), Wieczorkowski et al. (1999), Yee et al. (2015), and Al-Omari et al. (2019). Their proposals use estimators based on differences between order statistics, known as spacings.

Vasicek (1976) introduced one of the first non-parametric estimators based on spacings. Under the assumption that  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  is a random sample from the distribution  $F(z)$ , the estimator is defined as:

$$\hat{H}_V(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{n}{2m} (Z_{(i+m)} - Z_{(i-m)}) \right], \quad (3.9)$$

where  $m < n/2$  is a positive integer,  $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$  are the order statistics, and  $Z_{(i+m)} - Z_{(i-m)}$  is the  $m$ -spacing, in which  $Z_{(i)} = Z_{(1)}$  if  $i < 1$ ,  $Z_{(i)} = Z_{(n)}$  if  $i > n$ .

Several authors have explored adaptations to Vasicek's estimator. We consider the following entropy estimators variants, as discussed by Cassetti et al. (2022).

Van Es (1992) proposed a new estimator of entropy given by:

$$\hat{H}_{VE}(\mathbf{Z}) = \frac{1}{n-m} \sum_{i=1}^{n-m} \ln \left[ \frac{n+1}{m} (Z_{(i+m)} - Z_{(i)}) \right] + \sum_{k=m}^n \frac{1}{k} + \ln \frac{m}{n+1}. \quad (3.10)$$

Under some conditions, Van Es proved asymptotic normality of this estimator.

Ebrahimi et al. (1994) adjusted the weights of Vasicek's estimator, in order to take into account the fact that the differences are truncated around the smallest and the largest data points. Specifically,  $Z_{(i+m)} - Z_{(i-m)}$  is substituted with  $Z_{(i+m)} - Z_{(1)}$  when  $i \leq m$  and  $Z_{(i+m)} - Z_{(i-m)}$  is replaced by  $Z_{(n)} - Z_{(i-m)}$  when  $i \geq n - m + 1$ . Their estimator is given by:

$$\hat{H}_E(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{n}{c_i m} (Z_{(i+m)} - Z_{(i-m)}) \right], \quad (3.11)$$

where

$$c_i = \begin{cases} 1 + (i-1)/m & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 1 + (n-i)/m & \text{if } n-m+1 \leq i \leq n. \end{cases}$$

Correa (1995) suggested another modification of Vasicek's estimator. In estimation the density  $f$  of  $F$  in the interval  $(Z_{(i-m)}, Z_{(i+m)})$  he used a local linear model based on  $2m+1$  points:  $F(Z_{(j)}) = a + bZ_{(j)} + \varepsilon$ ,  $j = m-i, \dots, m+i$ . This yields a following estimator:

$$\hat{H}_C(\mathbf{Z}) = -\frac{1}{n} \sum_{i=1}^n \ln \frac{\sum_{j=i-m}^{i+m} (j-i) (Z_{(j)} - \bar{Z}_{(i)})}{n \sum_{j=i-m}^{i+m} (Z_{(j)} - \bar{Z}_{(i)})^2}, \quad (3.12)$$

where  $\bar{Z}_{(i)} = (2m+1)^{-1} \sum_{j=i-m}^{i+m} Z_{(j)}$ ,  $m < \frac{n}{2}$ ,  $Z_{(i)} = Z_{(1)}$  for  $i < 1$  and  $Z_{(i)} = Z_{(n)}$  for  $i > n$ . Based on simulations, he showed that his estimator has a smaller mean square error than Vasicek's approach.

Noughabi et al. (2010) modify the coefficients of (3.11) as:

$$\hat{H}_{NA}(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{n}{a_i m} (Z_{(i+m)} - Z_{(i-m)}) \right], \quad (3.13)$$

where

$$a_i = \begin{cases} 1 & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 1 & \text{if } n-m+1 \leq i \leq n, \end{cases}$$

and  $Z_{(i-m)} = Z_{(1)}$  for  $i \leq m$  and  $Z_{(i+m)} = Z_{(n)}$  for  $i \geq n-m$ .

Al-Omari (2014) suggested the following estimator:

$$\hat{H}_{AO}(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{n}{\omega_i m} (Z_{(i+m)} - Z_{(i-m)}) \right], \quad (3.14)$$

where

$$\omega_i = \begin{cases} 3/2 & \text{if } 1 \leq i \leq m, \\ 2 & \text{if } m+1 \leq i \leq n-m, \\ 3/2 & \text{if } n-m+1 \leq i \leq n, \end{cases}$$

in which  $Z_{(i-m)} = Z_{(1)}$  for  $i \leq m$ , and  $Z_{(i+m)} = Z_{(n)}$  for  $i \geq n - m$ .

These estimators are asymptotically consistent, i.e., they converge in probability to the true value when  $m, n \rightarrow \infty$  and  $m/n \rightarrow 0$ .

Our experimental setup involves an analysis of bias and mean squared error (MSE) for each estimator with a Monte Carlo study, using 1000 samples from the  $\Gamma_{\text{SAR}}$  distribution of size  $n \in \{9, 25, 49, 81, 121\}$ , with  $\mu \in \{1, 10, 50, 100\}$  and  $L = 5$ . The results are consistent with other situations. We use the heuristic spacing  $m = [\sqrt{n} + 0.5]$ , as recommended in the literature. Figure 3.3 presents the bias and MSE for each of the non-parametric entropy estimators. The results are summarized in Table 3.1.

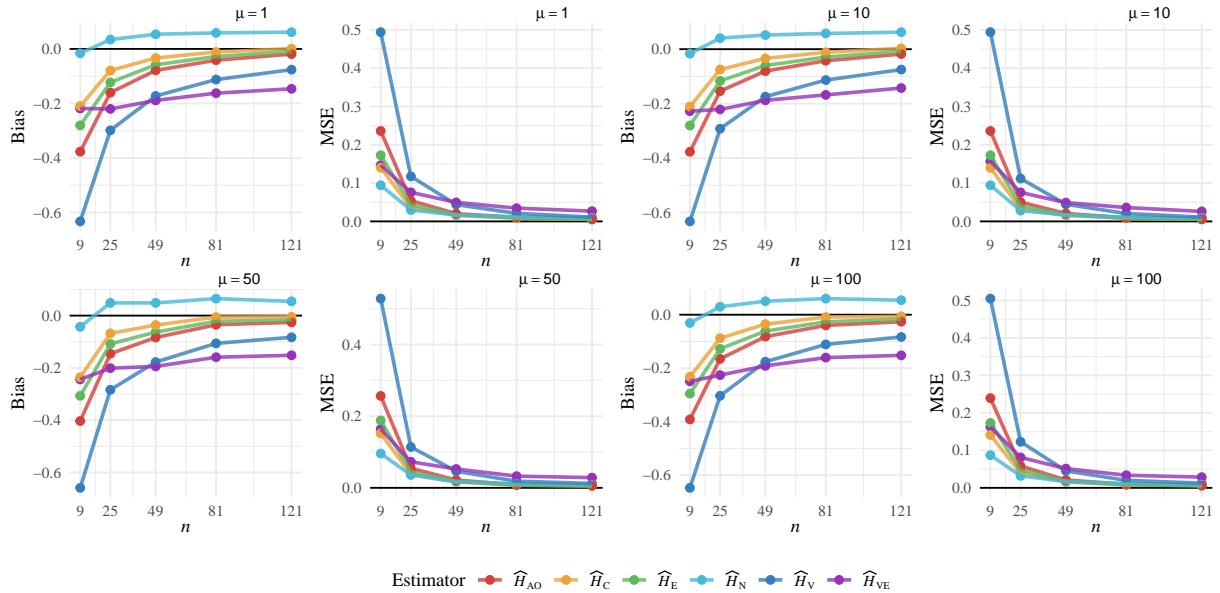


Figure 3.3: Bias and MSE of the entropy estimators for the  $\Gamma_{\text{SAR}}$ , with  $L = 5$ .

As shown in the simulation results, the estimators  $\widehat{H}_C$ ,  $\widehat{H}_E$ , and  $\widehat{H}_{AO}$  show low bias and achieve convergence for samples size larger than 81. These estimators exhibit good behavior in terms of bias and MSE across various parameter combinations.

In contrast, estimator  $\widehat{H}_N$  displays the lowest MSE across all scenarios. However, the bias remains unchanged and high for sample sizes larger than 25, indicating that the convergence is not very fast. Additionally, we observe that both  $\widehat{H}_V$  and  $\widehat{H}_{VE}$  estimators exhibit larger bias and slower convergence compared to their counterparts. Notably, the  $\widehat{H}_V$  estimator shows the highest MSE.

### 3.2.2 Enhanced estimators with Bootstrap

We extend our exploration of non-parametric entropy estimators by incorporating bootstrap methodologies, because we need them to perform well with small samples. The integration of bootstrap techniques aims to refine the accuracy of non-parametric entropy estimators.

Bootstrap is essentially a resampling algorithm introduced by (Efron, 1979). It offers a

Table 3.1: Bias and MSE of the entropy estimators for the  $\Gamma_{\text{SAR}}$ , with  $L = 5$ .

$\mu$	$n$	Bias						MSE					
		$\widehat{H}_V$	$\widehat{H}_{VE}$	$\widehat{H}_E$	$\widehat{H}_C$	$\widehat{H}_N$	$\widehat{H}_{AO}$	$\widehat{H}_V$	$\widehat{H}_{VE}$	$\widehat{H}_E$	$\widehat{H}_C$	$\widehat{H}_N$	$\widehat{H}_{AO}$
1	9	-0.632	-0.218	-0.280	-0.210	-0.016	-0.377	0.494	0.147	0.173	0.140	0.094	0.236
	25	-0.298	-0.220	-0.123	-0.079	0.034	-0.160	0.117	0.076	0.044	0.036	0.030	0.054
	49	-0.173	-0.189	-0.058	-0.033	0.054	-0.079	0.044	0.050	0.017	0.015	0.017	0.020
	81	-0.112	-0.162	-0.028	-0.011	0.059	-0.041	0.021	0.035	0.009	0.008	0.012	0.010
	121	-0.076	-0.146	-0.009	0.001	0.061	-0.019	0.011	0.027	0.005	0.005	0.009	0.006
10	9	-0.633	-0.228	-0.280	-0.211	-0.016	-0.377	0.494	0.157	0.173	0.140	0.094	0.236
	25	-0.292	-0.221	-0.117	-0.075	0.041	-0.154	0.112	0.076	0.040	0.032	0.028	0.050
	49	-0.174	-0.188	-0.060	-0.034	0.052	-0.080	0.045	0.049	0.018	0.016	0.017	0.021
	81	-0.113	-0.168	-0.029	-0.011	0.058	-0.042	0.021	0.036	0.009	0.008	0.011	0.009
	121	-0.075	-0.143	-0.008	0.003	0.063	-0.018	0.011	0.026	0.006	0.005	0.009	0.006
50	9	-0.659	-0.244	-0.307	-0.236	-0.043	-0.403	0.528	0.164	0.188	0.152	0.096	0.256
	25	-0.284	-0.201	-0.108	-0.068	0.049	-0.146	0.114	0.073	0.045	0.038	0.036	0.055
	49	-0.178	-0.194	-0.063	-0.036	0.049	-0.084	0.046	0.053	0.019	0.016	0.017	0.022
	81	-0.106	-0.159	-0.022	-0.006	0.065	-0.035	0.018	0.033	0.008	0.007	0.011	0.008
	121	-0.083	-0.152	-0.016	-0.004	0.054	-0.026	0.012	0.029	0.006	0.005	0.008	0.006
100	9	-0.647	-0.249	-0.295	-0.232	-0.031	-0.391	0.505	0.163	0.173	0.141	0.087	0.239
	25	-0.303	-0.226	-0.127	-0.088	0.030	-0.165	0.123	0.081	0.047	0.039	0.032	0.058
	49	-0.176	-0.191	-0.061	-0.035	0.050	-0.082	0.045	0.051	0.018	0.016	0.017	0.021
	81	-0.111	-0.160	-0.027	-0.009	0.060	-0.040	0.020	0.034	0.008	0.008	0.011	0.009
	121	-0.083	-0.152	-0.017	-0.006	0.054	-0.026	0.013	0.029	0.006	0.006	0.009	0.006

simulation-based approach for estimating various statistics of interest for a random variable. It can be used in both parametric and non-parametric settings. The main idea consists of creating new datasets from an existing one by resampling with repetition (Michelucci et al., 2021).

Let us assume that the non-parametric entropy estimator  $\widehat{H} = \widehat{\theta}(\mathbf{Z})$  is inherently biased, where  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  is a random sample from a distribution  $\mathcal{D}$ , i.e.:

$$\text{Bias}(\widehat{\theta}(\mathbf{Z})) = \mathbb{E}[\widehat{\theta}(\mathbf{Z})] - \theta \neq 0. \quad (3.15)$$

Our objective is to devise an unbiased estimator. To achieve this, we introduce an "ideal estimator"  $\check{\theta}(\mathbf{Z})$  using the bias information from (3.15):

$$\begin{aligned} \check{\theta}(\mathbf{Z}) &= \widehat{\theta}(\mathbf{Z}) - \text{Bias}(\widehat{\theta}(\mathbf{Z})), \\ &= \widehat{\theta}(\mathbf{Z}) - [\mathbb{E}[\widehat{\theta}(\mathbf{Z})] - \theta], \\ &= \widehat{\theta}(\mathbf{Z}) + \theta - \mathbb{E}[\widehat{\theta}(\mathbf{Z})]. \end{aligned} \quad (3.16)$$

However,  $\check{\theta}(\mathbf{Z})$  is not yet an estimator, because it depends on the true parameter, but  $\theta$  can be calculated using a non-parametric approach, and the average of individual estimators can be obtained using the bootstrap technique. This leads to the formulation of a new unbiased estimator, denoted as  $\widetilde{H}$ , from (3.16) we have:

$$\widetilde{H} = 2\widehat{\theta}(\mathbf{Z}) - \frac{1}{B} \sum_{b=1}^B \widehat{\theta}_b(\mathbf{Z}^{(b)}), \quad (3.17)$$

where  $B$  the number of bootstrap replications. For each iteration  $b$  (from 1 to  $B$ ), a sample  $\mathbf{Z}^{(b)}$  is generated, and the corresponding bootstrap estimator  $\widehat{\theta}_b(\mathbf{Z}^{(b)})$  is calculated. This methodology significantly reduces bias, resulting in more accurate and robust entropy estimations. The original estimators of Vasicek (3.9), Van Es (3.10), Ebrahimi (3.11), Correa (3.12), Noughabi (3.13), and Al-Omari (3.14) are now referred to as the proposed bootstrap-improved versions:  $\widetilde{H}_V$ ,  $\widetilde{H}_{VE}$ ,  $\widetilde{H}_E$ ,  $\widetilde{H}_C$ ,  $\widetilde{H}_{NA}$ , and  $\widetilde{H}_{AO}$ , respectively.

In order to assess the effectiveness of the bootstrap technique, we present comparisons of the bias and MSE between each original non-parametric entropy estimator and its respective bootstrap-enhanced version, with  $B = 200$  bootstrap samples, as shown in Figures 3.4 (a)–(f).

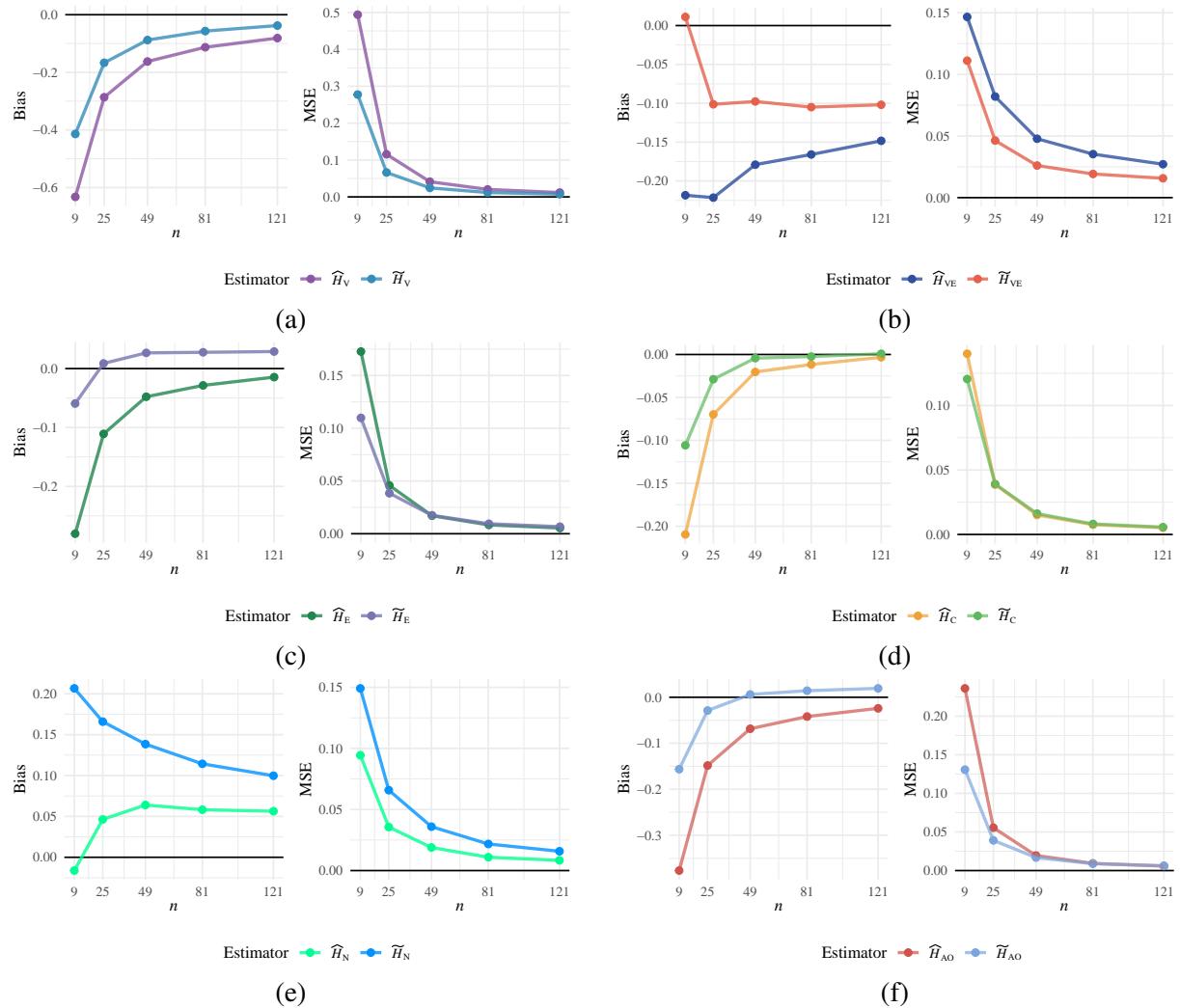


Figure 3.4: Comparing Bias and MSE: original vs. bootstrap estimators, with,  $\mu = 1$  and  $L = 5$ .

Based on previous simulations, the bootstrap technique did not improve the  $\widetilde{H}_{NA}$  estimator. This might occur because the original estimator overestimates the entropy values, showing a positive bias. This tendency to overestimate persists with the use of bootstrap, contributing to an increase in bias and MSE.

However, the improvement observed in the most estimators with the application of bootstrap techniques is notable, especially for sample sizes below 81. This performance is superior for the  $\tilde{H}_C$ ,  $\tilde{H}_E$ , and  $\tilde{H}_{AO}$  estimators. Therefore, we proceed to perform a comparison of bias and MSE of these three estimators alongside with their original versions, with 1000 samples from the  $\Gamma_{SAR}$  distribution of size  $n \in \{9, 25, 49, 81, 121\}$ , with  $\mu \in \{1, 10\}$ ,  $L = 5$ , and  $B = 200$  bootstrap samples, as shown in Figure 3.5 and Table 3.2.

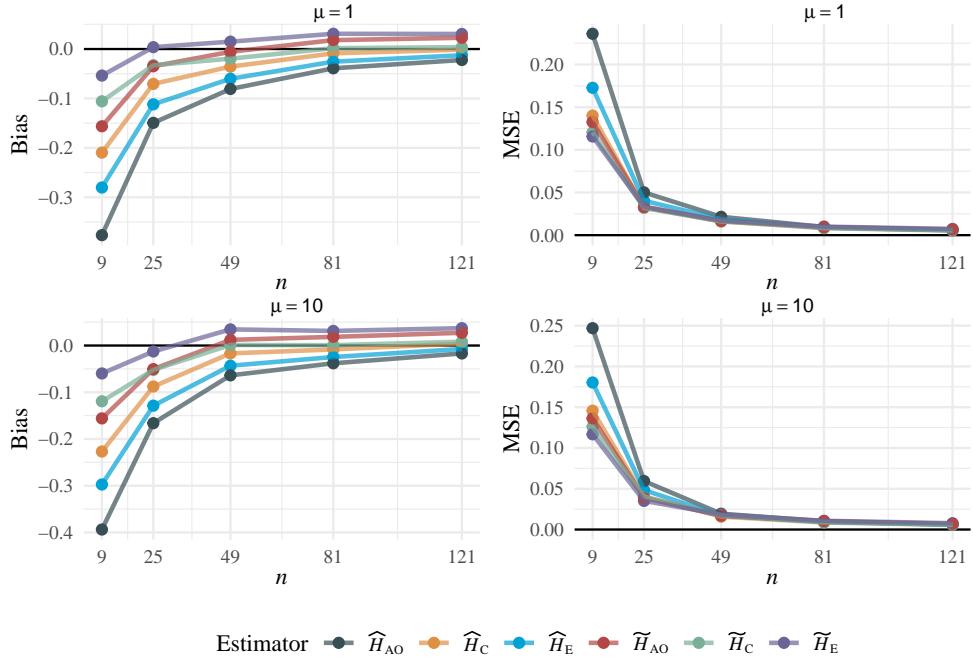


Figure 3.5: Bias and MSE of the entropy estimators for the  $\Gamma_{SAR}$ , with  $L = 5$ .

Table 3.2: Bias and MSE of the entropy estimators for the  $\Gamma_{SAR}$ , with  $L = 5$ .

$\mu$	$n$	Bias						MSE					
		$\widehat{H}_C$	$\widehat{H}_E$	$\widehat{H}_{AO}$	$\widetilde{H}_C$	$\widetilde{H}_E$	$\widetilde{H}_{AO}$	$\widehat{H}_C$	$\widehat{H}_E$	$\widehat{H}_{AO}$	$\widetilde{H}_C$	$\widetilde{H}_E$	$\widetilde{H}_{AO}$
1	9	-0.210	-0.280	-0.377	-0.106	-0.054	-0.156	0.140	0.173	0.236	0.121	0.116	0.133
	25	-0.071	-0.112	-0.149	-0.033	0.004	-0.035	0.032	0.040	0.050	0.032	0.033	0.033
	49	-0.036	-0.060	-0.081	-0.020	0.015	-0.006	0.016	0.019	0.021	0.016	0.017	0.017
	81	-0.009	-0.026	-0.039	0.002	0.031	0.018	0.008	0.009	0.010	0.009	0.010	0.009
10	121	-0.001	-0.013	-0.022	0.004	0.030	0.023	0.005	0.006	0.006	0.006	0.007	0.007
	9	-0.227	-0.297	-0.394	-0.119	-0.060	-0.156	0.146	0.180	0.247	0.126	0.117	0.136
	25	-0.088	-0.129	-0.166	-0.052	-0.013	-0.050	0.040	0.048	0.059	0.040	0.035	0.039
	49	-0.017	-0.043	-0.064	0.001	0.035	0.012	0.016	0.017	0.019	0.018	0.019	0.017
	81	-0.008	-0.025	-0.038	0.001	0.031	0.019	0.009	0.009	0.010	0.009	0.011	0.010
	121	0.004	-0.007	-0.017	0.008	0.037	0.027	0.006	0.006	0.006	0.006	0.008	0.007

### 3.3 Coefficient of variation and a robust alternative

The population CV is defined as a ratio of the population standard deviation ( $\sigma$ ) to the population mean ( $\mu$ ):

$$\text{CV} = \frac{\sigma}{\mu}, \quad \mu \neq 0. \quad (3.18)$$

We explore a robust alternative to CV, as described in (Ospina et al., 2019), which incorporates the ratio between the mean absolute deviation from the median (MnAD) and the median, two well-known robust measures of scale and location, respectively. The sample version for the MnAD is defined as  $n^{-1} \sum_{i=1}^n |x_i - \widehat{Q}_2|$ , where  $\widehat{Q}_2$  is an estimate for the median of the population, which can be approximated by the median of the sample.

### 3.4 Hypothesis Testing

We aim to test the following hypotheses:

$$\begin{cases} \mathcal{H}_0 : \text{The data come from the } \Gamma_{\text{SAR}} \text{ law,} \\ \mathcal{H}_1 : \text{The data come from the } G_I^0 \text{ distribution.} \end{cases}$$

We are testing the hypothesis that the data are fully-developed speckle versus the alternative of data with roughness. As for the parametric problem, once it is not possible to define the hypothesis  $\mathcal{H}_0 = \alpha = -\infty$ , it is impossible to solve this problem with parametric inference alternatives (such as likelihood ratio, score, gradient, and Wald hypothesis test). The proposed tests to solve this physical problem in SAR systems are described below.

#### 3.4.1 The Proposed Test Based on Non-parametric Entropy

For a random sample  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$  from a distribution  $\mathcal{D}$ , a test statistic is proposed. It is based on an empirical distribution that arises from the difference between non-parametrically estimated entropies  $\widetilde{H}(\mathbf{Z})$  and the analytical entropy of  $\Gamma_{\text{SAR}}$  (3.7) evaluated at the logarithm of the sample mean, where  $L \geq 1$  is known.

Hence, the entropy-based test statistic is defined as:

$$S(\mathbf{Z}; L) = \widetilde{H}(\mathbf{Z}) - [H_{\Gamma_{\text{SAR}}}(L) + \ln \overline{Z}]. \quad (3.19)$$

This test statistic aims to assess the behavior of the data under the null hypothesis using the empirical distribution. If the data represent fully-developed speckle, the density should center around zero, i.e.,  $S(\mathbf{Z}; L) \approx 0$ . Otherwise, the empirical distribution would shift from zero under the alternative hypothesis, suggesting significant differences and heterogeneous clutter.

The comparison between the bootstrap-improved estimators is shown in Table 3.3, where the test accuracy under the null hypothesis is presented alongside running times. The test accuracy is evaluated through 1000 simulated samples of different sizes, with each size replicated 100 times using bootstrap resampling.

The processing time is an important feature, especially considering the application of these estimators to large datasets of SAR images, as seen in Chapter 4.

Table 3.3: Test accuracy and processing time for each bootstrap-improved estimator.

<b>Estimator</b>	<b><math>L</math></b>	<b><math>n</math></b>	<b><math>S(\mathbf{Z}; L)</math></b>	<b>Time (s)</b>
$\tilde{H}_C$	2	25	-0.00152	22.53
		49	0.00515	40.35
		81	0.00625	63.93
		121	0.00751	97.06
	8	25	-0.04332	22.25
		49	-0.01659	33.42
		81	-0.00393	50.94
		121	0.00261	97.35
$\tilde{H}_E$	2	25	0.02204	4.66
		49	0.03452	5.55
		81	0.03195	6.89
		121	0.03012	7.90
	8	25	0.00801	4.81
		49	0.01654	5.43
		81	0.03036	6.38
		121	0.03137	7.46
$\tilde{H}_{AO}$	2	25	-0.01935	4.61
		49	0.00786	5.19
		81	0.01995	6.70
		121	0.01741	7.41
	8	25	-0.04020	4.74
		49	0.00047	5.35
		81	0.01176	6.21
		121	0.02019	7.48

As visible from Table 3.3, the accuracy of the test results across the three estimators shows similarities in specific sample sizes. However, practical scenarios in SAR image processing often involve small sample sizes, typically obtained over windows of size  $7 \times 7$ .

It is also noteworthy that the  $\tilde{H}_{AO}$  estimator exhibited the shortest processing time, followed by  $\tilde{H}_E$  and  $\tilde{H}_C$ . Considering this aspect, we select the  $\tilde{H}_{AO}$  estimator for subsequent simulations. Henceforth, the test statistical (3.19) will be denoted as:  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$ .

We now verify the normality of the data generated by the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  test. Figure 3.6 shows the empirical densities obtained by applying the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  test to different sample sizes drawn from the  $\Gamma_{SAR}$  distribution, where  $L$  takes values  $\{3, 5, 8, 11\}$  and  $\mu = 1$ .

Additionally, Table 3.4 summarizes the main descriptive statistics, including mean, standard deviation (SD), variance (Var), skewness (SK), excessive kurtosis (EK) and Anderson–Darling  $p$  values for normality.

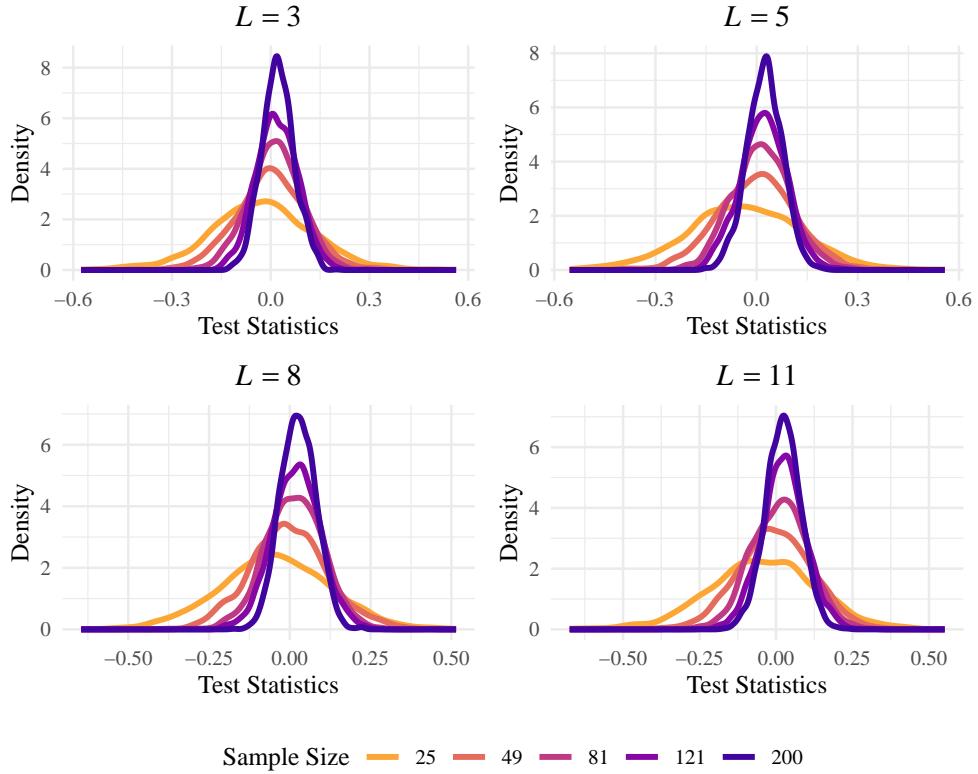


Figure 3.6: Empirical densities obtained from  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  test under the null hypothesis.

Table 3.4: Descriptive analysis of  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$ , with  $L \in \{3, 5, 8, 11\}$  and  $\mu = 1$ .

<b><math>L</math></b>	<b><math>n</math></b>	<b>Mean</b>	<b>SD</b>	<b>Var</b>	<b>SK</b>	<b>EK</b>	<b><math>p</math>-value</b>
3	25	-0.0280	0.1547	0.0239	-0.0076	0.4734	0.0022
	49	-0.0003	0.1053	0.0111	-0.0562	0.2610	0.0582
	81	0.0124	0.0796	0.0063	-0.0124	0.0536	0.5278
	121	0.0187	0.0630	0.0040	0.0337	-0.1826	0.5894
	200	0.0215	0.0490	0.0024	0.0625	-0.0473	0.0860
5	25	-0.0379	0.1669	0.0278	0.0007	0.1420	0.3267
	49	-0.0015	0.1150	0.0132	-0.1025	0.1998	0.2582
	81	0.0145	0.0869	0.0075	0.1008	0.5297	0.1121
	121	0.0198	0.0687	0.0047	0.0127	0.0222	0.2919
	200	0.0236	0.0529	0.0028	-0.0467	0.0977	0.3346
8	25	-0.0464	0.1680	0.0282	-0.0121	0.0980	0.6477
	49	-0.0031	0.1202	0.0144	0.1282	0.2000	0.0038
	81	0.0137	0.0883	0.0078	-0.0279	0.2554	0.5567
	121	0.0200	0.0738	0.0055	-0.0089	0.0686	0.7502
	200	0.0260	0.0546	0.0030	0.0716	-0.0349	0.3771
11	25	-0.0442	0.1735	0.0301	-0.1422	0.2413	0.0981
	49	-0.0019	0.1201	0.0144	-0.0503	0.1464	0.9576
	81	0.0127	0.0917	0.0084	-0.0172	0.0333	0.3179
	121	0.0239	0.0729	0.0053	-0.0127	0.2102	0.0596
	200	0.0234	0.0572	0.0033	-0.0233	0.1072	0.6740

Results with  $p$  values greater than 0.05 do not indicate a violation of the normality assumption. A low variance, skewness, and excessive kurtosis of almost zero indicate limited dispersion, asymmetry, and a light tail.

Normal Q–Q plots confirm no evidence against a normal distribution, as shown in Figure 3.7.

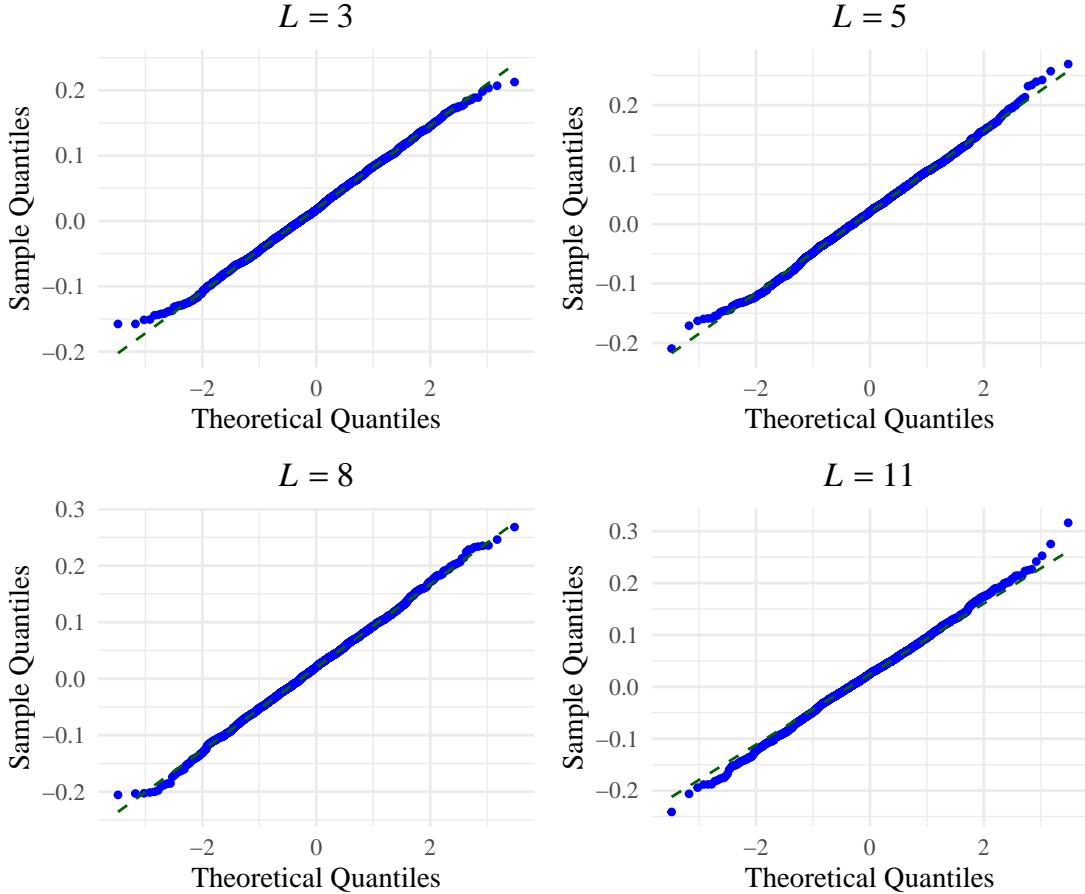


Figure 3.7: Normal Q–Q plots for  $n = 121$ .

After checking the data's normality, we examined the proposed test's abilities in terms of size and power. Under  $\mathcal{H}_0$ , the distribution of the test statistic is asymptotically normal. Therefore, the  $p$  values are calculated as  $2\Phi(-|\varepsilon|)$ , where  $\Phi$  is the standard Gaussian cumulative distribution function, and  $\varepsilon$  is the standardized test statistic given by:

$$\varepsilon = \frac{\tilde{H}_{AO}(\mathbf{Z}) - [H_{\Gamma_{SAR}}(L) + \ln \bar{Z}]}{\hat{\sigma}}.$$

We have nominal levels of 1 %, 5 %, and 10 %. In terms of size, 1000 simulations were used for different sample sizes from the  $\Gamma_{SAR}$  distribution, with varying values of  $L$ , and  $\mu = 1$ . In all cases, the nominal level was achieved. We assessed the test power using 1000 simulations for different sample sizes from the  $G_I^0$  distribution, with  $\mu = 1$ , and  $\alpha = -2$ . The power generally improves with increasing sample size and number of looks. The results are shown in Table 3.5.

Table 3.5: Size and Power of the  $S_{\tilde{H}_{AO}}(\mathbf{Z})$  test statistic.

$L$	$n$	Size			Power		
		1%	5%	10%	1%	5%	10%
3	25	0.0160	0.0620	0.1070	0.6900	0.8450	0.8340
	49	0.0100	0.0480	0.0960	0.6890	0.8920	0.8480
	81	0.0120	0.0490	0.1080	0.6260	0.8750	0.8540
	121	0.0090	0.0690	0.1190	0.5680	0.8620	0.8230
5	25	0.0210	0.0660	0.1130	0.9120	0.9620	0.9880
	49	0.0100	0.0460	0.1080	0.9470	0.9820	0.9960
	81	0.0120	0.0560	0.1070	0.9580	0.9900	0.9960
	121	0.0150	0.0640	0.1150	0.9420	0.9780	0.9950
8	25	0.0210	0.0650	0.1080	0.9930	0.9950	0.9970
	49	0.0060	0.0470	0.0860	0.9980	1.0000	0.9970
	81	0.0120	0.0490	0.1000	0.9930	0.9980	0.9990
	121	0.0150	0.0650	0.1220	0.9970	0.9990	0.9980
11	25	0.0130	0.0610	0.1000	0.9990	0.9990	0.9990
	49	0.0100	0.0450	0.0920	0.9980	0.9990	0.9990
	81	0.0170	0.0530	0.1050	1.0000	1.0000	1.0000
	121	0.0160	0.0680	0.1180	0.9980	1.0000	0.9980

### 3.4.2 The Proposed Test Based on Coefficient of Variation and a Robust Alternative

In addition to the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  test, we also propose a test statistic based on the classical CV. This test statistic is defined as follows:

$$T_{CV} = \frac{S}{\bar{Z}}, \quad (3.20)$$

where  $S$  and  $\bar{Z}$  are the sample standard deviation and mean, respectively.

Similarly, we use another test statistic based on the ratio of the MnAD to the median. This statistic is given by:

$$T_{CV_{MnAD}} = \frac{\text{MnAD}}{\text{Median}}. \quad (3.21)$$

We proceed to identify suitable models for these estimators of the CV, and then form test statistics.

The situations in which the use of CV and  $CV_{MnAD}$  may be appropriate, i.e., when the observations are positive, the log-normal (LN) and the inverse Gaussian distribution (IG) are often more appropriate than the Gamma and Weibull distributions (Chaubey et al., 2017; Takagi et al., 1997).

It is shown that the IG distribution is well approximated by the log-normal distribution, which means that the IG distribution also does not share the problem of the non-existence of a fixed-width confidence interval with the Gaussian case (Whitmore et al., 1978).

The biparametric LN distribution has density:

$$f_Z(z; \mu_{LN}, \sigma_{LN}) = \frac{1}{\sigma_{LN} z \sqrt{2\pi}} \exp \left\{ -\frac{(\ln z - \mu_{LN})^2}{2\sigma_{LN}^2} \right\} \mathbb{1}_{\mathbb{R}_+}(z), \quad (3.22)$$

with  $\mu_{LN}$  is any real number, and  $\sigma_{LN}$  is positive.

### 3.4.3 Model Selection Criterion

We used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to select the best-fitting distribution.

The AIC deals with the trade-off between the goodness-of-fit and the model's simplicity in terms of the number of model parameters (Burnham et al., 2004). The model or distribution with the lowest value of AIC is chosen to be the best. The BIC assesses goodness-of-fit of a distribution or model, but avoids overfitting by penalising additional degrees of freedom (Dziak et al., 2019). The model with the lowest BIC value is chosen as the best.

The AIC and BIC results in Tables 3.6–3.9 indicate that the CV and  $CV_{MnAD}$  data from different distributed  $\Gamma_{SAR}$  and  $G_I^0$  synthetic sample sizes match the properties of an LN distribution. It is important to note that this conclusion was drawn empirically based on a dictionary of analytically tractable distributions and well-defined under biparametric, unimodal, asymmetric, and positive distributions.

Figures 3.9–3.11 show empirical and fitted density plots, Q–Q plots, P–P plots, as well as empirical and fitted cumulative distribution functions. They provide qualitative sources that confirm that the LN distribution is the most appropriate distribution. As expected, both test statistics work well under the null hypothesis. Under the alternative hypothesis, the P–P plot shows that  $CV_{MnAD}$  is more robust than CV, although both statistics suffer from the tail effect caused by the distributed  $G_I^0$  data.

Table 3.6: AIC and BIC values for evaluating the best distribution with CV data from  $\Gamma_{SAR}$ .

<b>Criterion</b>	<b><i>n</i></b>	<b>Normal</b>	<b>Lognormal</b>	<b>Gamma</b>	<b>Weibull</b>	<b>Inverse Gaussian</b>
AIC	25	−38031.9	−38266.7	−38311.8	−36413.2	−38261.6
	49	−47698.2	−47913.7	−47905.6	−45554.0	−47911.6
	81	−55382.1	−55494.7	−55494.9	−53220.4	−55493.8
	121	−61344.9	−61470.8	−61453.8	−58876.0	−61470.5
BIC	25	−38016.7	−38251.5	−38296.6	−36398.0	−38246.4
	49	−47683.0	−47898.5	−47890.4	−45538.7	−47896.4
	81	−55366.9	−55479.5	−55479.6	−53205.2	−55478.6
	121	−61329.7	−61455.6	−61438.6	−58860.8	−61455.2

Table 3.7: AIC and BIC values for evaluating the best distribution with CV data from  $G_I^0$ .

<b>Criterion</b>	<b><i>n</i></b>	<b>Normal</b>	<b>Lognormal</b>	<b>Gamma</b>	<b>Weibull</b>	<b>Inverse Gaussian</b>
AIC	25	8254.04	2186.31	3628.40	8383.58	2257.63
	49	8821.79	1689.03	3483.10	9533.53	1835.29
	81	8525.81	866.29	2853.31	9822.48	1057.91
	121	8708.81	131.86	2341.06	10506.49	398.53
BIC	25	8269.27	2201.54	3643.63	8398.81	2272.86
	49	8837.02	1704.26	3498.33	9548.76	1850.52
	81	8541.04	881.52	2868.55	9837.72	1073.14
	121	8724.04	147.09	2356.29	10521.72	413.76

Table 3.8: AIC and BIC values for evaluating the best distribution with  $\text{CV}_{\text{MnAD}}$  data from  $\Gamma_{\text{SAR}}$ .

<b>Criterion</b>	<b><i>n</i></b>	<b>Normal</b>	<b>Lognormal</b>	<b>Gamma</b>	<b>Weibull</b>	<b>Inverse Gaussian</b>
AIC	25	-38375.56	-39147.85	-39066.29	-36652.49	-39143.61
	49	-48386.11	-48795.32	-48745.48	-46240.91	-48793.83
	81	-56072.87	-56322.32	-56290.02	-53836.04	-56322.12
	121	-62217.14	-62394.77	-62369.32	-59861.80	-62394.57
BIC	25	-38360.32	-39132.62	-39051.05	-36637.26	-39128.38
	49	-48370.87	-48780.09	-48730.25	-46225.67	-48778.60
	81	-56057.64	-56307.09	-56274.79	-53820.81	-56306.89
	121	-62201.91	-62379.54	-62354.09	-59846.56	-62379.33

Table 3.9: AIC and BIC values for evaluating the best distribution with  $\text{CV}_{\text{MnAD}}$  data from  $G_I^0$ .

<b>Criterion</b>	<b><i>n</i></b>	<b>Normal</b>	<b>Lognormal</b>	<b>Gamma</b>	<b>Weibull</b>	<b>Inverse Gaussian</b>
AIC	25	-13302.39	-15575.23	-15158.42	-11933.35	-15565.03
	49	-23265.27	-24529.72	-24284.47	-20986.94	-24522.28
	81	-29908.19	-30960.38	-30747.90	-25233.41	-30946.54
	121	-36496.78	-37128.07	-36991.41	-32366.52	-37123.20
BIC	25	-13287.16	-15559.99	-15143.19	-11918.12	-15549.80
	49	-23250.04	-24514.48	-24269.24	-20971.71	-24507.05
	81	-29892.96	-30945.15	-30732.67	-25218.17	-30931.31
	121	-36481.55	-37112.84	-36976.18	-32351.28	-37107.97

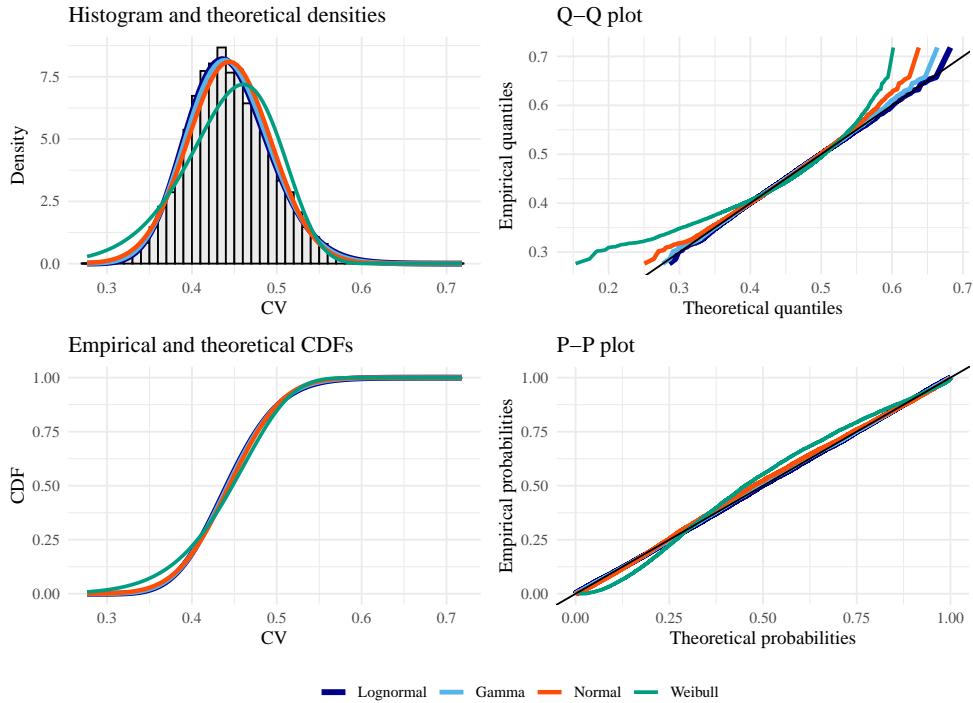


Figure 3.8: Goodness of fit plots for evaluating the best distribution with CV data from  $\Gamma_{\text{SAR}}$  (under the null hypothesis), with  $n = 49$ ,  $L = 5$ , and  $\mu = 1$ .

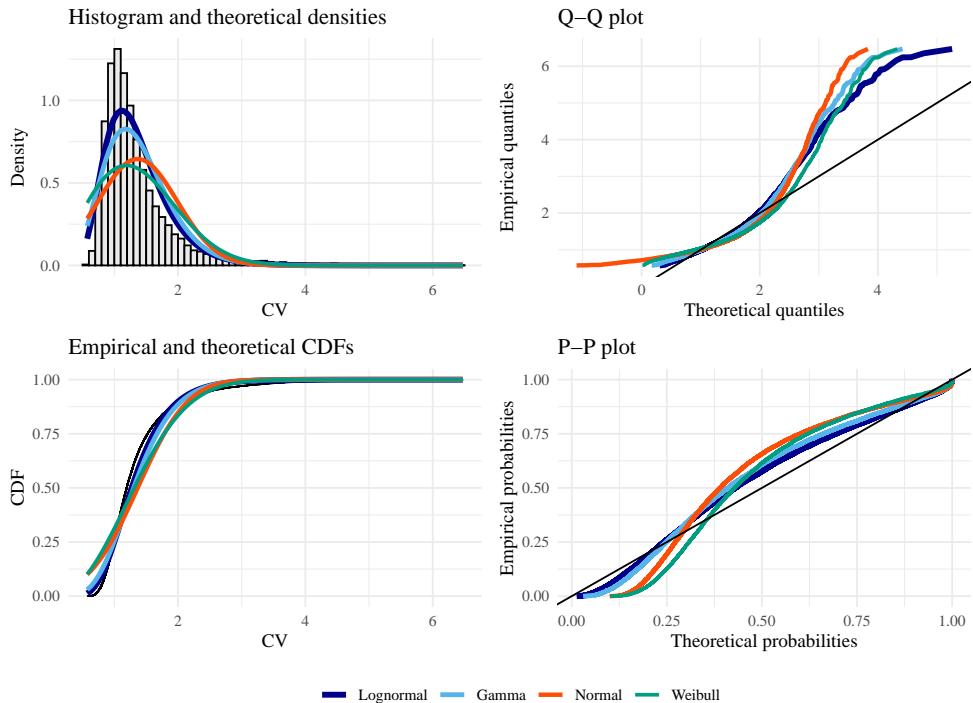


Figure 3.9: Goodness of fit plots for evaluating the best distribution with CV data from  $G_I^0$  (under the alternative hypothesis), with  $n = 49$ ,  $L = 5$ ,  $\mu = 1$ , and  $\alpha = -3$ .

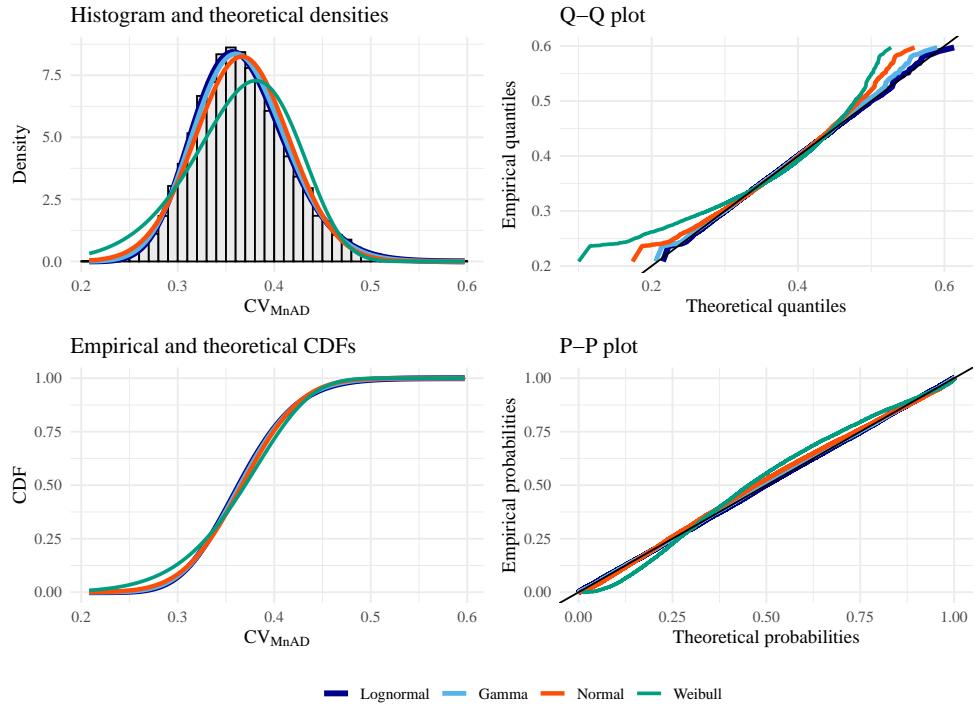


Figure 3.10: Goodness of fit plots for evaluating the best distribution with  $\text{CV}_{\text{MnAD}}$  data from  $\Gamma_{\text{SAR}}$  (under the null hypothesis), with  $n = 49$ ,  $L = 5$ , and  $\mu = 1$ .

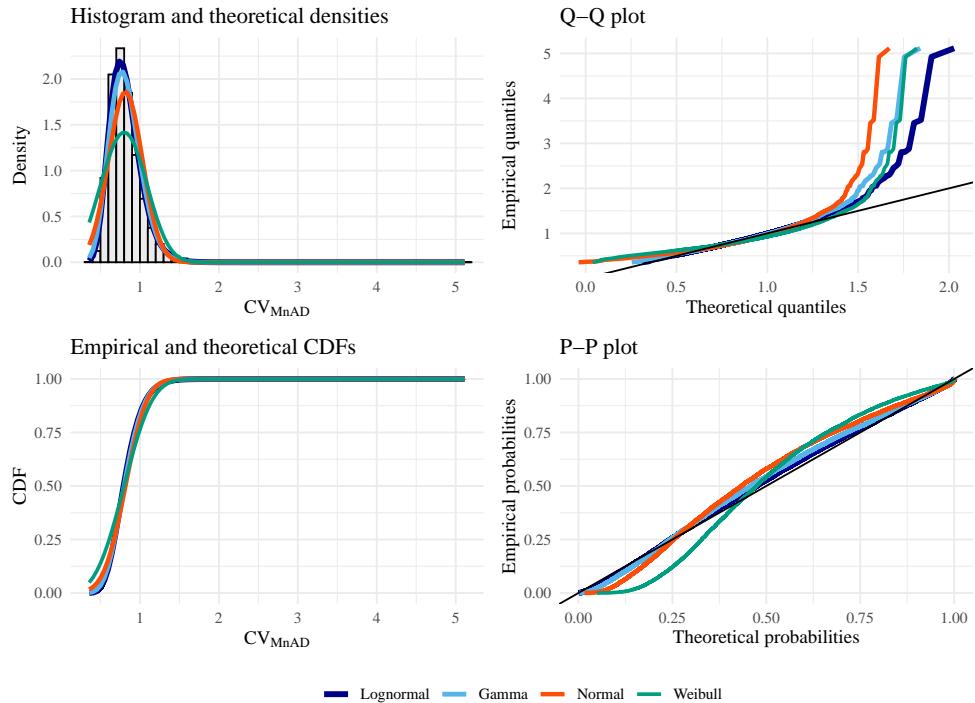


Figure 3.11: Goodness of fit plots for evaluating the best distribution with  $\text{CV}_{\text{MnAD}}$  data from  $G_I^0$  (under the alternative hypothesis), with  $n = 49$ ,  $L = 5$ ,  $\mu = 1$ , and  $\alpha = -3$ .

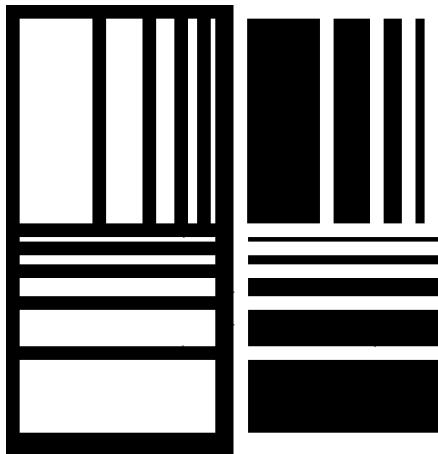
# Results

This section presents the simulations we performed to evaluate the proposed test statistics' performance, followed by applications to SAR data.

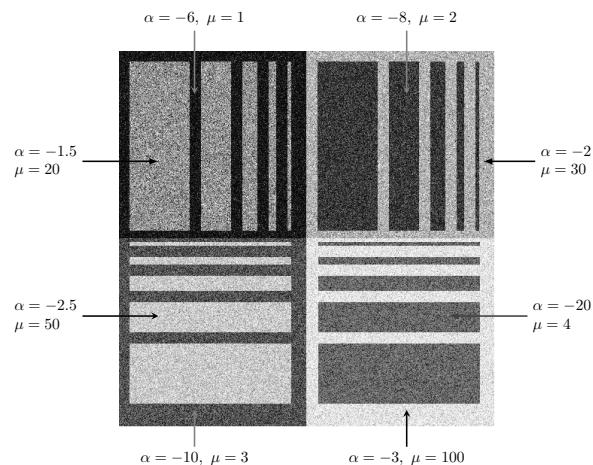
## 4.1 Simulated Data

Figure 4.1 (a) shows the phantom with dimensions of  $500 \times 500$  pixels. It was proposed by Gomez et al. (2017) as a tool to assess the performance of speckle-reduction filters.

Figure 4.1 (b) shows the simulated image, where each small phantom displaying texture variations. The observations are independent draws from the  $G_I^0$  distribution (3.5), with  $L = 5$  and varying  $\alpha$  and  $\mu$ , annotated in the image for each quadrant. Light regions correspond to textured observations (heterogeneous), while darker regions represent textureless areas (homogeneous).



(a) Phantom.



(b) Simulated image, varying  $\alpha$  and  $\mu$ , with  $L = 5$ .

Figure 4.1: Synthetic dataset.

The  $\alpha$  parameter of the  $G_I^0$  distribution is essential for interpreting texture characteristics.

Values near zero greater than  $-3$  suggest extremely textured targets, such as urban zones (Frery et al., 2019). As the value decreases, it indicates regions with moderate texture (in the  $[-6, -3]$  region), related to forest zones, while values below  $-6$  correspond to textureless regions, such as pasture, agricultural fields, and water bodies (Neto et al., 2023).

We applied the three test statistics, namely  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$ ,  $T_{CV}$ , and  $T_{CV_{MnAD}}$ , to the simulated image using local sliding windows of size  $7 \times 7$ , as shown in Figures 4.2 (a)–(c).

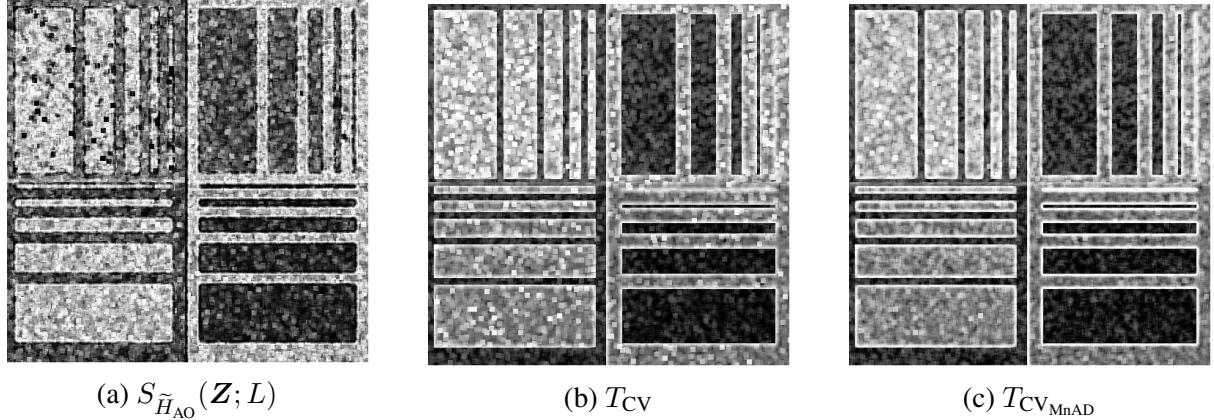


Figure 4.2: Results of applying the test statistics to the simulated image.

The resulting  $p$ -values for each test are shown in Figures 4.3 (a)–(c). In Figures 4.4 (a)–(c), maps are depicted using a color table between black, gray levels, and white. All  $p$ -values above 0.05 are represented in white (indicating no evidence to reject the null hypothesis), while those below 0.05 are shown in black (indicating evidence to reject the hypothesis). We notice that the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  performs significantly better than the other tests in identifying heterogeneous areas in the simulated image.

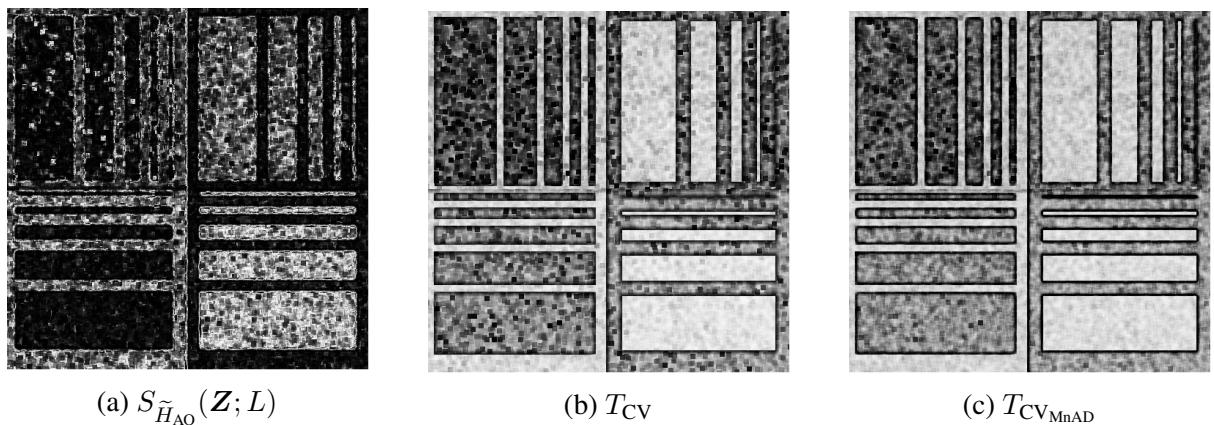


Figure 4.3: Map of  $p$ -values of simulated image for each test.

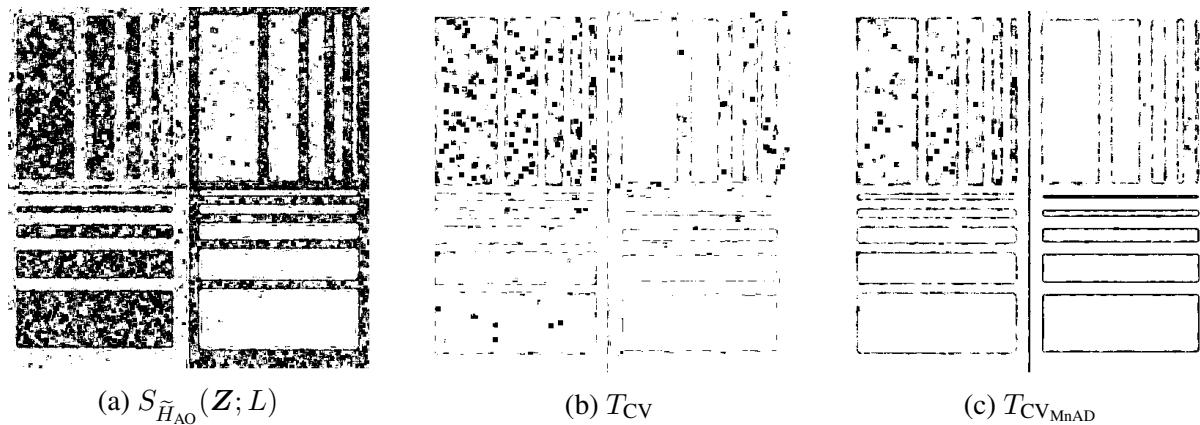
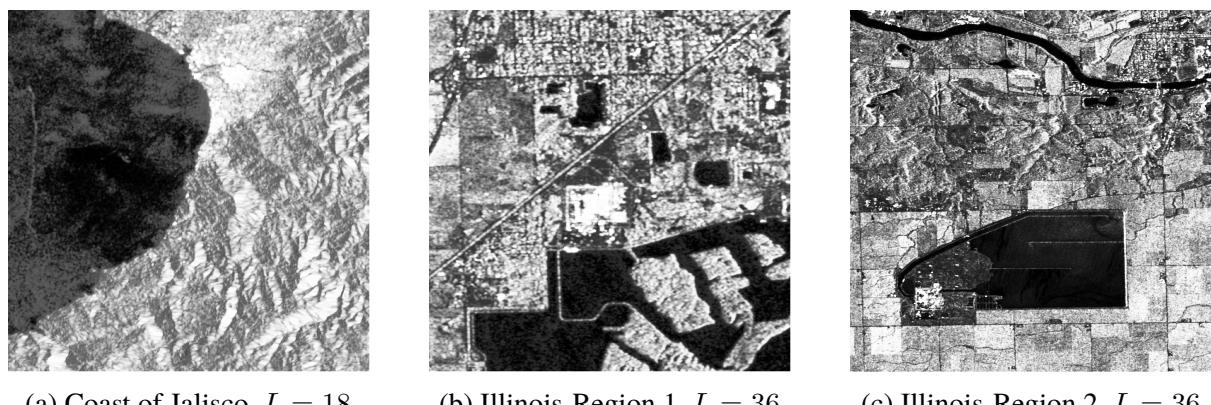


Figure 4.4: Results for a threshold of 0.05 of the  $p$ -value of simulated image for each test.

## 4.2 SAR Data

We evaluated the proposed test statistics using three SAR images: one of the coast of Jalisco, Mexico (with a spatial resolution of 20 m both along azimuth and range directions) and two of Illinois, USA (with a spatial resolution of 10 m both along azimuth and range directions), acquired by the Sentinel-1B satellite operating in C-band, with VV polarization and intensity format. The first two images have a size of  $512 \times 512$  pixels, while the third has  $1024 \times 1024$  pixels, and they contain mountainous areas, agricultural regions, water bodies, and urban areas, as shown in Figures 4.5 (a)–(c).



(a) Coast of Jalisco,  $L = 18$

(b) Illinois-Region 1,  $L = 36$

(c) Illinois-Region 2,  $L = 36$

Figure 4.5: SAR images.

The three statistical tests are applied to the SAR images using  $7 \times 7$  local sliding windows, as illustrated in Figures 4.6, 4.9 and 4.12.

The  $p$ -values obtained for each test are presented in Figures 4.7, 4.10 and 4.13, respectively.

In Figures 4.8, 4.11 and 4.14, the maps of  $p$ -values composed of a linear gradient of black and white colors, represent the decisions at a 5 % significance level. Dark areas represent values below 0.05, indicating evidence to reject the null hypothesis and suggesting heterogeneity in these

regions. In contrast, values above 0.05 are represented as white areas, indicating no evidence to reject the fully-developed speckle hypothesis.

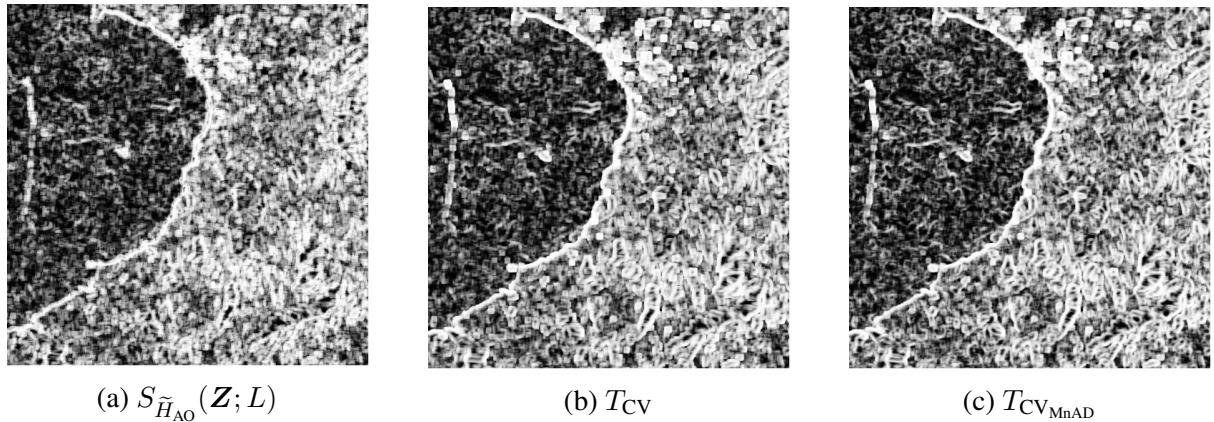


Figure 4.6: Results of applying the test statistics to Coast of Jalisco image.

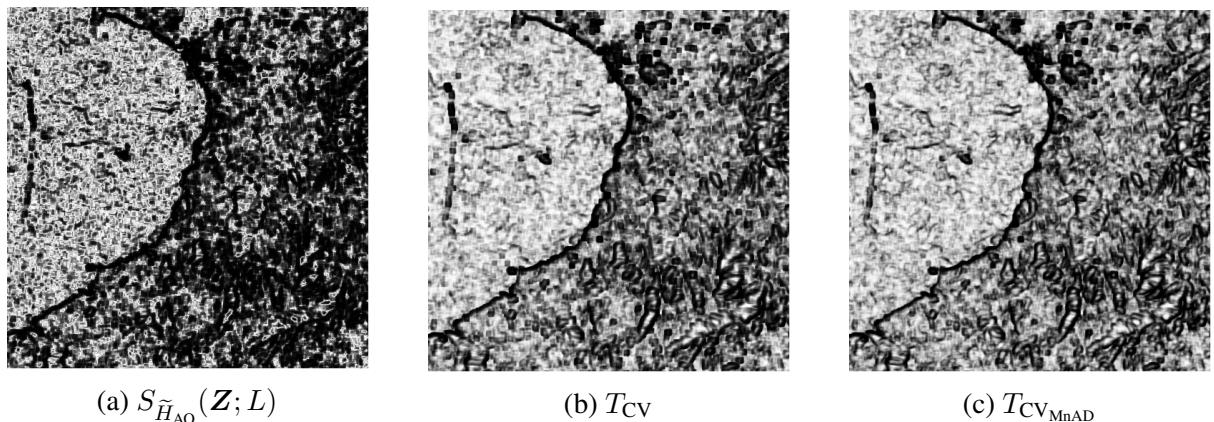


Figure 4.7: Map of  $p$ -values of Coast of Jalisco image for each test.

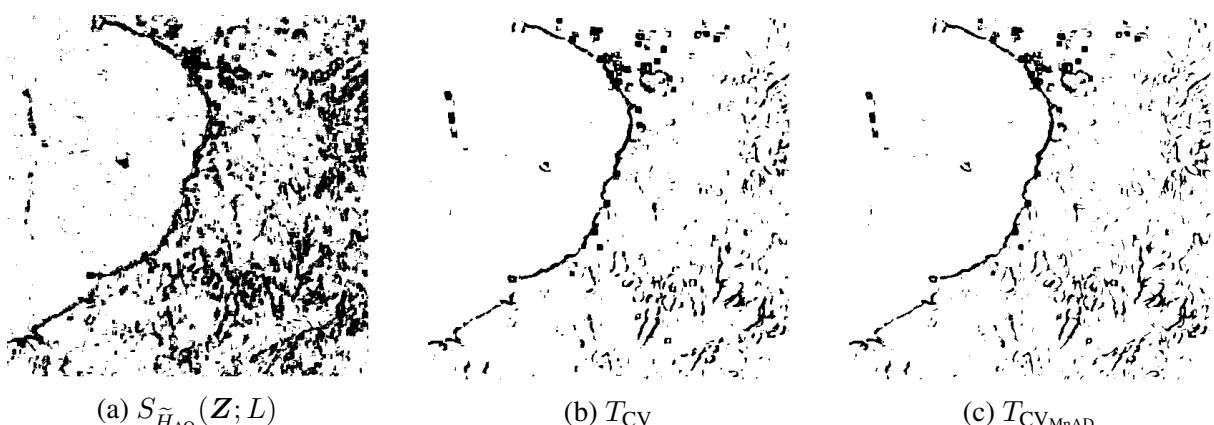


Figure 4.8: Results for a threshold of 0.05 of the  $p$ -value of Coast of Jalisco for each test.

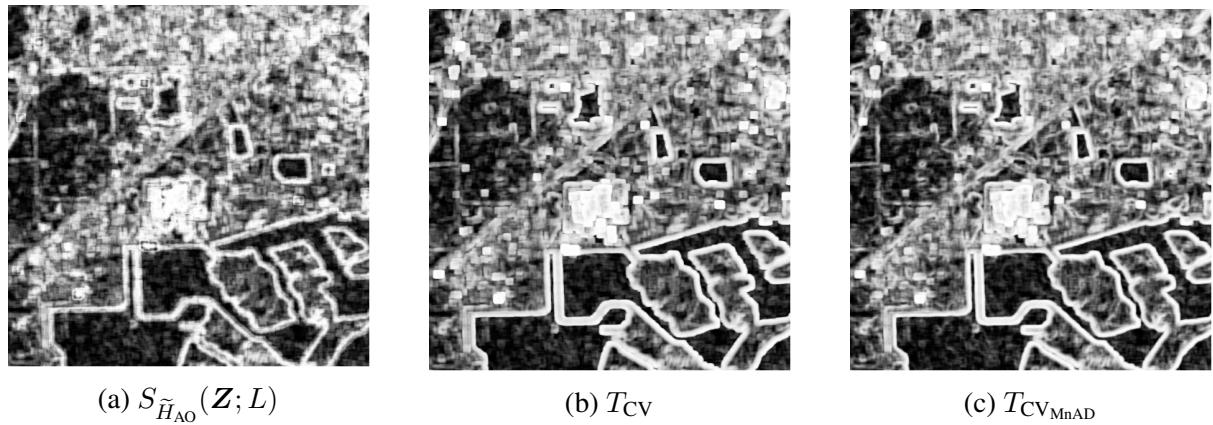


Figure 4.9: Results of applying the test statistics, Illinois-Region 1.

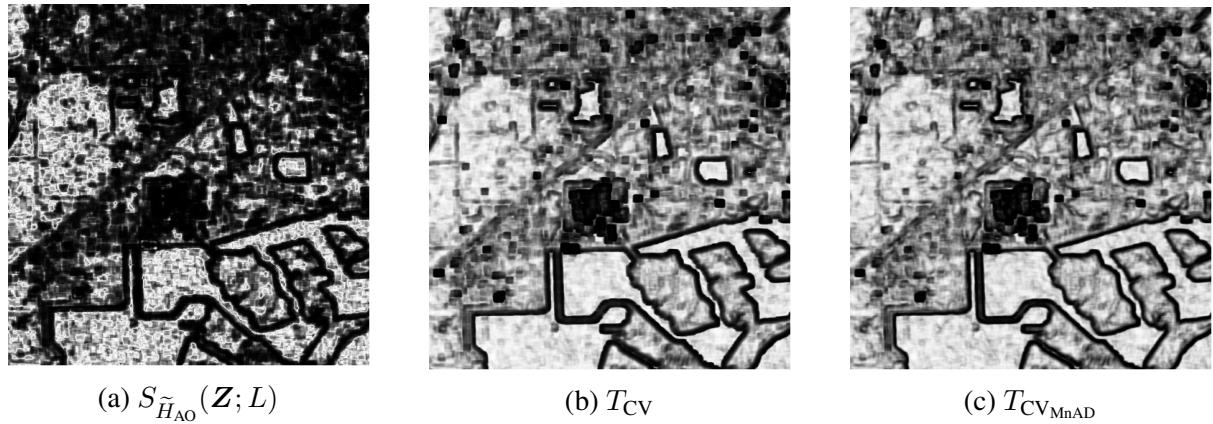


Figure 4.10: Map of  $p$ -values, Illinois-Region 1.

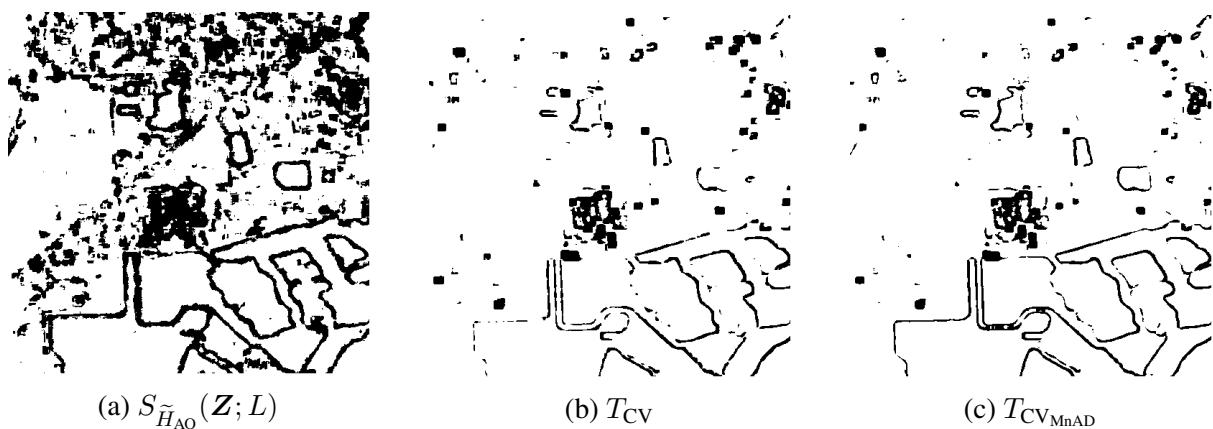


Figure 4.11: Results for a threshold of 0.05 of the  $p$ -value, Illinois-Region 1.

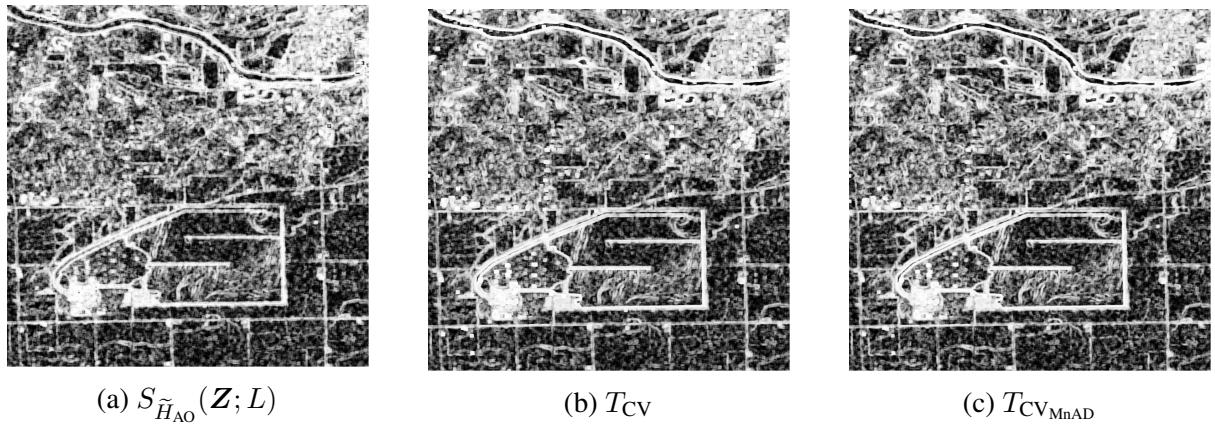


Figure 4.12: Results of applying the test statistics, Illinois-Region 2.

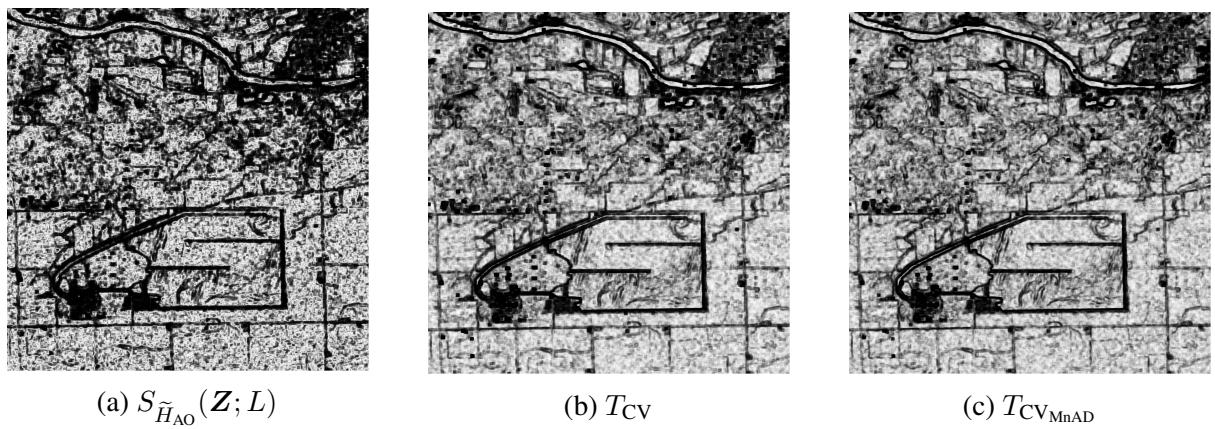


Figure 4.13: Map of  $p$ -values, Illinois-Region 2.

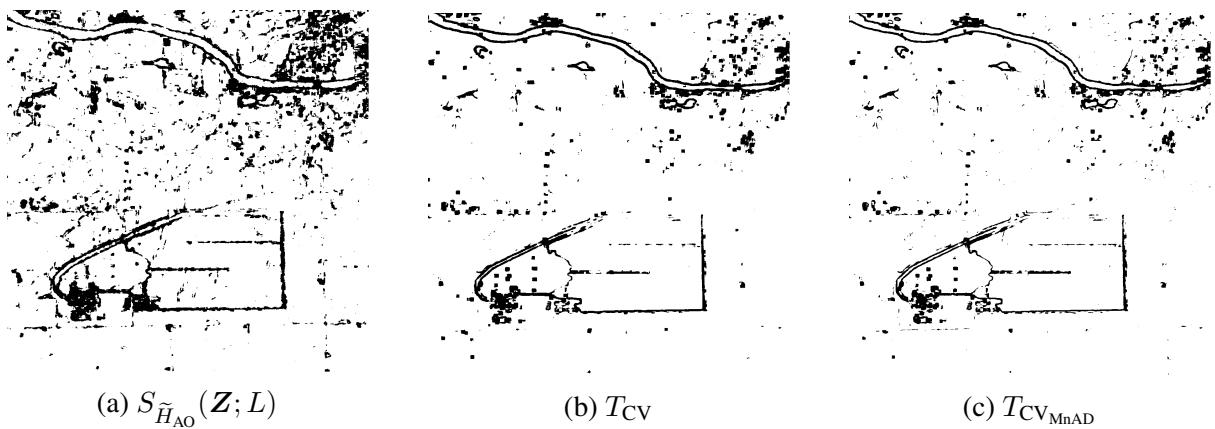


Figure 4.14: Results for a threshold of 0.05 of the  $p$ -value, Illinois-Region 2.

Using Shannon entropy is more meaningful than using the original and robust CV to capture heterogeneity. It is justified that the dark areas of the maps based on the  $T_{CV}$  and  $T_{CV_{MnAD}}$  show

coverage patterns similar to those reported for the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  map. This suggests that although CV-based tests may produce slightly less pronounced results than the entropy-based test, they still demonstrate a comparable ability to detect heterogeneity within SAR images.

It is noticeable that the entropy and CV-based tools predicted heterogeneity regions and boundaries where the statistical properties of texture vary. The  $T_{CV_{MnAD}}$  test was shown to be an effective edge detector. It emerges as a robust alternative to the classical CV test, making it less susceptible to the influence of outliers and allowing it to produce more precise edges. Considering a higher significance level may increase the sensitivity to edge detection but also increase the risk of detecting false heterogeneous regions.

Additionally, assuming a 5 % threshold for  $p$ -values, in most cases, the heterogeneous regions detected by the  $S_{\tilde{H}_{AO}}(\mathbf{Z}; L)$  test were more extensive than those detected by the  $T_{CV}$  and  $T_{CV_{MnAD}}$  tests. This was mainly observable in Figures 4.4 (a), 4.8 (a), and 4.11 (a).

# Conclusions and future perspectives

This work provides a practical and theoretical answer to the following physical question: How to detect heterogeneity in SAR images, assuming that the SAR intensity follows the  $\Gamma_{\text{SAR}}$  model. To this end, we proposed three novel hypothesis tests, one from the Shannon entropy and two from the variation coefficient variants. The performance of our proposals was evaluated using a Monte Carlo study. The results showed that they were conservative in estimating the probability of a type I error (false alarm rate) and the test power (probability of detection), which increases with sample size. An application to three recent SAR images was performed. The results showed that the Shannon entropy-based test was more robust than the CV-based tests. In addition, all tests could recognize images with different textures and identify edges where the texture type changes.

## Future perspectives

For future research directions, there is a promising avenue in estimating the equivalent number of looks (ENL), a crucial parameter in the statistical modeling of multi-look synthetic aperture radar (SAR) imagery. This parameter serves as an indicator of heterogeneity and can significantly impact the accuracy of statistical analyses.

Additionally, we plan to explore the estimation of ENL in polarimetric SAR (PolSAR) data. This extension will involve:

- Estimating the test on each of the three intensity channels of fully PolSAR data
- Analyzing the joint distribution
- Proposing techniques that generalize the test statistic into the analysis of PolSAR data. First, we can examine the entropy and CV of the distribution for the eigenvalues of the coherence matrix and then the Shannon entropy for the PolSAR matrix and trace-based versions for the CV of the PolSAR matrix.

# Bibliography

- Al-Omari, A. I. (2014). “Estimation of entropy using random sampling”. In: *Journal of Computational and Applied Mathematics* 261, pp. 95–102 (cit. on p. 15).
- Al-Omari, A. I. and Haq, A. (2019). “Novel entropy estimators of a continuous random variable”. In: *International Journal of Modeling, Simulation, and Scientific Computing* 10.02, p. 1950004 (cit. on p. 14).
- Amitrano, D., Di Martino, G., Guida, R., Iervolino, P., Iodice, A., Papa, M. N., Riccio, D., and Ruello, G. (2021). “Earth Environmental Monitoring Using Multi-Temporal Synthetic Aperture Radar: A Critical Review of Selected Applications”. In: *Remote Sensing* 13.4, p. 604 (cit. on p. 1).
- Argenti, F., Lapini, A., Bianchi, T., and Alparone, L. (2013). “A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images”. In: *IEEE Geoscience and Remote Sensing Magazine* 1.3, pp. 6–35 (cit. on p. 1).
- Avval, T. G., Moeini, B., Carver, V., Fairley, N., Smith, E. F., Baltrusaitis, J., Fernandez, V., Tyler, B. J., Gallagher, N., and Linford, M. R. (2021). “The Often-Overlooked Power of Summary Statistics in Exploratory Data Analysis: Comparison of Pattern Recognition Entropy (PRE) to Other Summary Statistics and Introduction of Divided Spectrum-PRE (DS-PRE)”. In: *Journal of Chemical Information and Modeling* 61.9, pp. 4173–4189 (cit. on p. 1).
- Banik, S. and Kibria, B. M. G. (2011). “Estimating the Population Coefficient of Variation by Confidence Intervals”. In: *Communications in Statistics - Simulation and Computation* 40.8, pp. 1236–1261 (cit. on p. 2).
- Burnham, K. P. and Anderson, D. R. (2004). “Multimodel Inference: Understanding AIC and BIC in Model Selection”. In: *Sociological Methods & Research* 33.2, pp. 261–304 (cit. on p. 25).
- Campbell, J. B. and Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford press (cit. on p. 4).
- Cassetti, J., Delgadino, D., Rey, A., and Frery, A. C. (2022). “Entropy Estimators in SAR Image Classification”. In: *Entropy* 24.4, p. 509 (cit. on pp. 12, 14).

- Chankham, W., Niwitpong, S.-A., and Niwitpong, S. (2024). “The Simultaneous Confidence Interval for the Ratios of the Coefficients of Variation of Multiple Inverse Gaussian Distributions and Its Application to PM2.5 Data”. In: *Symmetry* 16.3, p. 331 (cit. on p. 2).
- Chaubey, Y. P., Singh, M., and Sen, D. (2017). “Symmetrizing and Variance Stabilizing Transformations of Sample Coefficient of Variation from Inverse Gaussian Distribution”. In: *Sankhya B* 79.2, pp. 217–246 (cit. on p. 24).
- Correa, J. C. (1995). “A new estimator of entropy”. In: *Communications in Statistics-Theory and Methods* 24.10, pp. 2439–2449 (cit. on pp. 14, 15).
- Cumming, I. G. and Wong, F. H. (2005). “Digital processing of synthetic aperture radar data”. In: *Artech house* 1.3, pp. 108–110 (cit. on pp. 5, 9).
- De A. Ferreira, J. and Nascimento, A. D. C. (2020). “Shannon Entropy for the  $\mathcal{G}_I^0$  Model: A New Segmentation Approach”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13, pp. 2547–2553 (cit. on p. 1).
- Dingle Robertson, L., M. Davidson, A., McNairn, H., Hosseini, M., Mitchell, S., Abelleyra, D. de, Verón, S., Maire, G. le, Plannells, M., Valero, S., Ahmadian, N., Coffin, A., Bosch, D., H. Cosh, M., Basso, B., and Saliendra, N. (2020). “C-band synthetic aperture radar SAR imagery for the classification of diverse cropping systems”. In: *International Journal of Remote Sensing* 41.24, pp. 9628–9649 (cit. on p. 1).
- Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., and Jermiin, L. S. (2019). “Sensitivity and specificity of information criteria”. In: *Briefings in Bioinformatics* 21.2, pp. 553–565 (cit. on p. 25).
- Ebrahimi, N., Pflughoeft, K., and Soofi, E. S. (1994). “Two measures of sample entropy”. In: *Statistics & Probability Letters* 20.3, pp. 225–234 (cit. on p. 15).
- Efron, B. (1979). “Bootstrap Methods: Another Look at the Jackknife”. In: *The Annals of Statistics* 7.1 (cit. on p. 16).
- Everaerts, J (2005). “PEGASUS—Bridging the gap between airborne and spaceborne remote sensing”. In: *New Strategies For European Remote Sensing*, pp. 395–401 (cit. on p. 5).
- Frery, A. C., Cribari-Neto, F., and Souza, M. O. (2004). “Analysis of Minute Features in Speckled Imagery with Maximum Likelihood Estimation”. In: *EURASIP Journal on Advances in Signal Processing* 2004.16, pp. 2476–2491 (cit. on p. 1).
- Frery, A. C., Muller, H.-J., Yanasse, C., and Sant'Anna, S. (1997). “A model for extremely heterogeneous clutter”. In: *IEEE Transactions on Geoscience and Remote Sensing* 35.3, pp. 648–659 (cit. on pp. 10, 12).
- Frery, A. C. and Gambini, J. (2019). “Comparing samples from the  $\mathcal{G}^0$  distribution using a geodesic distance”. In: *TEST* 29.2, pp. 359–378 (cit. on p. 30).
- Gomez, L., Alvarez, L., Mazorra, L., and Frery, A. C. (2017). “Fully PolSAR image classification using machine learning techniques and reaction-diffusion systems”. In: *Neurocomputing* 255, pp. 52–60 (cit. on p. 29).
- Hendricks, K. W. R. (1936). “The Sampling Distribution of the Coefficient of Variation”. In: *The Annals of Mathematical Statistics, Vol. 7, No. 3*, pp. 129-132 (cit. on p. 2).

- Hughes, L. H. (2020). “Deep learning for matching high-resolution SAR and optical imagery”. PhD thesis. Technische Universität München (cit. on p. 9).
- Michelucci, U. and Venturini, F. (2021). “Estimating Neural Network’s Performance with Bootstrap: A Tutorial”. In: *Machine Learning and Knowledge Extraction* 3.2, pp. 357–373 (cit. on p. 17).
- Mohammad-Djafari, A. (2015). “Entropy, Information Theory, Information Geometry and Bayesian Inference in Data, Signal and Image Processing and Inverse Problems”. In: *Entropy* 17.6, pp. 3989–4027 (cit. on p. 1).
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., and Papathanassiou, K. P. (2013). “A tutorial on synthetic aperture radar”. In: *IEEE Geoscience and Remote Sensing Magazine* 1.1, pp. 6–43 (cit. on p. 6).
- Mu, Y., Huang, B., Pan, Z., Yang, H., Hou, G., and Duan, J. (2019). “An Enhanced High-Order Variational Model Based on Speckle Noise Removal With  $G^0$  Distribution”. In: *IEEE Access* 7, pp. 104365–104379 (cit. on p. 1).
- Nascimento, A. D. C., Cintra, R. J., and Frery, A. C. (2010). “Hypothesis Testing in Speckled Data With Stochastic Distances”. In: *IEEE Transactions on Geoscience and Remote Sensing* 48.1, pp. 373–385 (cit. on p. 12).
- Nascimento, A. D. C., Frery, A. C., and Cintra, R. J. (2019). “Detecting Changes in Fully Polarimetric SAR Imagery With Statistical Information Theory”. In: *IEEE Transactions on Geoscience and Remote Sensing* 57.3, pp. 1380–1392 (cit. on p. 1).
- Nascimento, A. D. C., Horta, M. M., Frery, A. C., and Cintra, R. J. (2014). “Comparing Edge Detection Methods Based on Stochastic Entropies and Distances for PolSAR Imagery”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.2, pp. 648–663 (cit. on p. 1).
- Neto, J. F. S. R. and Rodrigues, F. A. A. (2023). “Improving Log-Cumulant-Based Estimation of Heterogeneity Information in SAR Imagery”. In: *IEEE Geoscience and Remote Sensing Letters* 20, pp. 1–5 (cit. on p. 30).
- Noughabi, A. (2010). “A new estimator of entropy and its application in testing normality”. In: *Journal of Statistical Computation and Simulation* 80.10, pp. 1151–1162 (cit. on p. 2).
- Noughabi, A. and Arghami, N. (2010). “A New Estimator of Entropy”. In: *Journal of the Iranian Statistical Society* 9.1, pp. 53–64 (cit. on p. 15).
- Oliver, C. and Quegan, S. (2004). *Understanding synthetic aperture radar images*. SciTech Publishing (cit. on p. 9).
- Orth, P. H. R. (2018). *The Geometry of Spaceborne Synthetic Aperture Radar* (cit. on p. 8).
- Ospina, R. and Marmolejo-Ramos, F. (2019). “Performance of Some Estimators of Relative Variability”. In: *Frontiers in Applied Mathematics and Statistics* 5 (cit. on p. 20).
- Pearson, K. and Henrici, O. M. F. E. (1896). “VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia”. In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 187, pp. 253–318 (cit. on p. 2).

- Pressé, S., Ghosh, K., Lee, J., and Dill, K. A. (2013). “Principles of maximum entropy and maximum caliber in statistical physics”. In: *Reviews of Modern Physics* 85.3, pp. 1115–1141 (cit. on p. 1).
- Richards, M. A. (2005). *Fundamentals of radar signal processing*. Vol. 1. Mcgraw-hill New York (cit. on p. 4).
- Sara, D., Mandava, A. K., Kumar, A., Duela, S., and Jude, A. (2021). “Hyperspectral and multispectral image fusion techniques for high resolution applications: a review”. In: *Earth Science Informatics* 14.4, pp. 1685–1705 (cit. on p. 5).
- Shannon, C. E. (1948). “A Mathematical Theory of Communication”. In: *The Bell System Technical Journal* 27.3, pp. 379–423 (cit. on p. 1).
- Subhash, S., Sunoj, S. M., Sankaran, P. G., and Rajesh, G. (2021). “Nonparametric estimation of quantile-based entropy function”. In: *Communications in Statistics-Simulation and Computation* 52.5, pp. 1805–1821 (cit. on p. 2).
- Subrahmanya, K. and Aruna Rao, K. (2003). “Tests of coefficients of variation of normal population”. In: *Communications in Statistics-Simulation and Computation* 32.3, pp. 641–661 (cit. on p. 2).
- Takagi, K., Kumagai, S., Matsunaga, I., and Kusaka, Y. (1997). “Application of inverse Gaussian distribution to occupational exposure data”. In: *The Annals of Occupational Hygiene* 41.5, pp. 505–514 (cit. on p. 24).
- Tian, L. (2005). “Inferences on the common coefficient of variation”. In: *Statistics in Medicine* 24.14, pp. 2213–2220 (cit. on p. 2).
- Touzi, R., Lopes, A., and Bousquet, P. (1988). “A statistical and geometrical edge detector for SAR images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 26.6, pp. 764–773 (cit. on p. 2).
- Ulaby, F., Kouyate, F., Brisco, B., and Williams, T. H. (1986). “Textural Information in SAR Images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 24.2, pp. 235–245 (cit. on p. 2).
- Van Es, B. (1992). “Estimating Functionals Related to a Density by a Class of Statistics Based on Spacings”. In: *Scandinavian Journal of Statistics* 19, pp. 61–72 (cit. on p. 14).
- Vasconcellos, K. L. P., Frery, A. C., and Silva, L. B. (2005). “Improving Estimation in Speckled Imagery”. In: *Computational Statistics* 20.3, pp. 503–519 (cit. on p. 1).
- Vasicek, O. (1976). “A test for normality based on sample entropy”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 38.1, pp. 54–59 (cit. on p. 14).
- Wang, J., Zheng, T., Lei, P., and Bai, X. (2019). “A hierarchical convolution neural network (CNN)-based ship target detection method in spaceborne SAR imagery”. In: *Remote Sensing* 11.6, p. 620 (cit. on p. 1).
- Whitmore, G. and Yalovsky, M (1978). “A normalizing logarithmic transformation for inverse Gaussian random variables”. In: *Technometrics* 20.2, pp. 207–208 (cit. on p. 25).
- Wieczorkowski, R. and Grzegorzewski, P. (1999). “Entropy estimators-improvements and comparisons”. In: *Communications in Statistics-Simulation and Computation* 28.2, pp. 541–567 (cit. on pp. 2, 14).

- Wiley, C. A. (1985). "Synthetic aperture radars". In: *IEEE Transactions on Aerospace and Electronic Systems* 3, pp. 440–443 (cit. on p. 4).
- Yee, J., Kwon, M.-S., Jin, S., Park, T., and Park, M. (2015). "Detecting Genetic Interactions for Quantitative Traits Using m-Spacing Entropy Measure". In: *BioMed Research International* 2015, pp. 1–10 (cit. on p. 14).
- Yue, D.-X., Xu, F., Frery, A. C., and Jin, Y.-Q. (2021). "Synthetic Aperture Radar Image Statistical Modeling: Part One-Single-Pixel Statistical Models". In: *IEEE Geoscience and Remote Sensing Magazine* 9.1, pp. 82–114 (cit. on p. 1).
- Zamanzade, E. and Arghami, N. R. (2012). "Testing normality based on new entropy estimators". In: *Journal of Statistical Computation and Simulation* 82.11, pp. 1701–1713 (cit. on p. 2).
- Zhang, D. (2022). "Synthetic Aperture Radar Image interpretation based on deep learning". PhD thesis. University of Hamburg (cit. on p. 6).
- Zhao, X., Shao, W., Lai, Z., and Jiang, X. (2023). "Retrieval of Rain Rates for Tropical Cyclones From Sentinel-1 Synthetic Aperture Radar Images". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16, pp. 3187–3197 (cit. on p. 1).