

Entropy Estimators in SAR Image Classification

Julia Cassetti, Daiana Delgadino, Andrea Rey and Alejandro C. Frery, *Senior Member*

Abstract—Remotely sensed data are successfully used for information extraction. In particular SAR imagery, which suffers the presence of speckle noise, needs special models and techniques. In this sense, the \mathcal{G}^0 family of distributions is a suitable model for SAR intensity because it can characterize areas with different degrees of texture. Information theory has gained a place in signal and image processing for parameter estimation and feature extraction. Among its tools, entropy stands out as one of the most expressive features. In this paper, we evaluate the performance of several parametric and non parametric Shannon entropy estimators as input for supervised and unsupervised classification algorithms for SAR images. These classification algorithms were analyzed through appropriate index. Finally we apply these methodology to actual data.

Index Terms—Feature extraction, synthetic aperture radar, Shannon entropy estimator, classification.

I. INTRODUCTION

IMAGES that come from coherent illumination systems, such as those acquired by synthetic aperture radar (SAR) are contaminated by speckle noise. This kind of noise corrupts the image making it difficult to analyze and interpret.

Under this perspective, statistical procedures are important tools for processing SAR data. The choice of a suitable model and appropriate measures to describe this sort of images is a fundamental point to obtain features that promote a good analysis. In this sense, the family of distributions \mathcal{G}^0 [1] has been extensively used to model SAR data because of its ability to a wide variety of roughness targets.

Several approaches have been developed in order to obtain expressive and tractable features. In particular, entropy measures have been widely used for this purpose. Estimation parameter [2], classification [3], methodologies for constructing confidence interval and contrast measures [4, 5], and edge detection [6] are some examples of its application.

Sundry authors have tackled the segmentation and classification SAR images problem using information theory measures. Nobre et al. [7] used Rényi's entropy for monopolarized SAR image segmentation. Ferreira et al. [8] derived a closed-form expression for the Shannon entropy based on the \mathcal{G}^0 law for intensity data, and proposed a new entropy-based segmentation method. Carvalho et al. [3] employed stochastic distances to approach unsupervised classification methodology applied to Polarimetric Synthetic Aperture Radar (PolSAR) images. Palacio et al. [9] used machine learning techniques in

Julia Cassetti jcassetti@campus.uns.edu.ar and Daiana Delgadino ddelgadino@campus.uns.edu.ar are with the Instituto de Desarrollo Humano, Universidad Nacional de General Sarmiento, Pcia. de Buenos Aires, Argentina.

Andrea Rey arey@frba.utn.edu.ar is with Mathematics Department and Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina.

Alejandro C. Frery alejandro.frery@vuw.ac.nz is with the School of Mathematics and Statistics, Victoria University of Wellington, New Zealand.

combination with filters to perform a classification in PolSAR images.

The Shannon entropy has been applied to analyzed SAR imagery in several approaches, from inference [4] to classification [8]. Therefore, its estimation deserves attention. Vasicek [10] replaced the distribution function F by the empirical distribution function F_n and used a difference operator in place of the differential operator in the Shannon entropy expression. Van Es [11] studied an entropy estimator based on differences between order statistics. Correa [12] proposed a new entropy estimator determined from local linear regression. Al-Omari [13] and Noughabi and Noughabi [14] presented modified versions of the estimator introduced by Ebrahimi et al. [15].

In this paper, we address the classification problem through supervised and unsupervised strategies whose input is entropy local estimation. To this aim, we evaluate entropy estimators in both cases, parametric and non-parametric. In the parametric case we use the relationship between \mathcal{G}^0 and Fisher distributions to obtain an expression of the entropy. In the non-parametric case, we assess these estimators in terms of bias, mean square error, computational time and accuracy when classifying a monopolarimetric SAR image.

II. THE \mathcal{G}^0 MODEL

The multiplicative model defines the return Z in a monopolarized SAR image as the product of two independent random variables: one corresponding to the backscatter X , and the other one to the speckle noise Y . In this manner, $Z = XY$ represents the return in each pixel of the image.

The \mathcal{G}^0 distribution family becomes attractive because of its flexibility to adequately model areas with all types of roughness [16, 17]. For intensity SAR data, this family arises from considering the speckle noise Y modeled as a Γ distributed random variable with unitary mean and shape parameter $L \geq 1$, the number of looks, while the backscatter X is considered to obey a Reciprocal of Gamma law. Thus, the density function for intensity data is given by

$$f(z) = \frac{L^L \Gamma(L - \alpha)}{\gamma^\alpha \Gamma(-\alpha) \Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}}, \quad (1)$$

where $-\alpha, \gamma, z > 0$ and $L \geq 1$. The r -order moment is

$$\mathbb{E}(Z^r) = \left(\frac{\gamma}{L}\right)^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)} \cdot \frac{\Gamma(L + r)}{\Gamma(L)}, \quad (2)$$

provided $\alpha < -r$, and infinite otherwise.

Mejail et al. [16] proved a relationship between \mathcal{G}^0 distribution and the Fisher-Snedekor F law, which states that the cumulative distribution function $F_{\alpha, \gamma, L}$ for the return Z is

$$F_{\alpha, \gamma, L}(z) = \Upsilon_{2L, -2\alpha}(-\alpha z / \gamma), \quad (3)$$

for every $z > 0$, where $\Upsilon_{2L, -2\alpha}$ is the cumulative distribution function of a Fisher-Snedekor random variable with $2L$ and -2α degrees of freedom. This connection is useful to obtain a close entropy formula.

III. SHANNON ENTROPY

It is well known Shannon's contribution to the creation of what is known as Information Theory. Shannon [18] proposed a new way of measuring the transmission of information through a channel, thinking information as a statistical concept. In this regard, Shanon entropy (SE) is defined as

$$H[f(x)] = -E[\log f(x)] = - \int_{-\infty}^{\infty} f(x) \log f(x) dx, \quad (4)$$

where X is a continuous random variable with probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$. The entropy of the \mathcal{G}^0 distribution can be obtained using (3). Denote H_F the entropy under the F model, then the \mathcal{G}^0 entropy for intensity data $H_{\mathcal{G}^0}$ is

$$H_{\mathcal{G}^0}(\alpha, \gamma, L) = H_F(2L, -2\alpha) - \log(-\alpha/\gamma). \quad (5)$$

Using (5), the expression of $H_{\mathcal{G}^0}$ is

$$\begin{aligned} H_{\mathcal{G}^0}(\alpha, \gamma, L) &= -\log(-\alpha/\gamma) - (1-\alpha)\psi^{(0)}(-\alpha) \\ &\quad + \log(-\alpha/L) + (L-\alpha)\psi^{(0)}(L-\alpha) \\ &\quad + \log(B(L, -\alpha)) + (1-L)\psi^{(0)}(L), \end{aligned} \quad (6)$$

where $\psi^{(0)}$ and B are the digamma and beta functions, respectively.

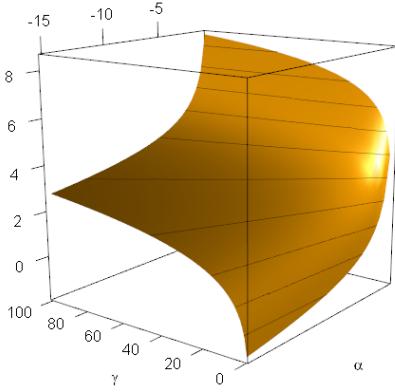


Fig. 1. $H_{\mathcal{G}^0}(\alpha, \gamma, L)$ as a function of alpha and gamma for $L = 3$.

Figure 1 shows the $H_{\mathcal{G}^0}(\alpha, \gamma, L)$ theoretical entropy as a function of α and γ with $L = 3$. It can be shown that, for each fixed γ value, $H_{\mathcal{G}^0}$ is an injective function. The same behavior repeats if we consider α as a constant.

IV. SHANNON ENTROPY ESTIMATORS

Several authors have proposed entropy estimators according to (4). Most of them are based on order statistics of the sample. Al-Omari [13] presented an overview of these estimators and also proposed a new one. From a parametric point of view, is natural to consider the maximum likelihood estimator (MV) of the entropy (HMV).

In what follows, the entropy estimators studied in this paper are described.

A. Maximum likelihood entropy estimator

Let Z_1, \dots, Z_n be an independent random sample of size n from the $\mathcal{G}^0(\alpha, \gamma, L)$ distribution. Assume L is known. The maximum likelihood estimator of α and γ for L known, denoted $\hat{\alpha}_{\text{ML}}$ and $\hat{\gamma}_{\text{ML}}$ respectively, consists of the values in the parametric space $\mathbb{R}_- \times \mathbb{R}_+$, that maximize the loglikelihood function:

$$\begin{aligned} &\log \Gamma(L - \hat{\alpha}_{\text{ML}}) - \hat{\alpha}_{\text{ML}} \log \hat{\gamma}_{\text{ML}} - \log \Gamma(-\hat{\alpha}_{\text{ML}}) \\ &\quad + \frac{\hat{\alpha}_{\text{ML}} - L}{n} \sum_{i=1}^n \log (\hat{\gamma}_{\text{ML}} + LZ_i). \end{aligned} \quad (7)$$

Optimal asymptotic properties of the MV estimator are well-known and its calculation demands numerical maximization routines. We use the L-BFGS-B version of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [19] that allows box constraints. This algorithm belongs to the family of quasi-Newton methods, that do not require the Hessian matrix, but only the gradient.

The ML entropy estimator [20] is

$$H_{\text{ML}} = H_{\mathcal{G}^0}(\hat{\alpha}_{\text{MV}}, \hat{\gamma}_{\text{MV}}, L). \quad (8)$$

B. Non-parametric entropy estimators

The problem of estimating entropy has attracted the interest of several authors. For continuous random variables Vasicek [10] showed that the functional defined in (4) can be written as

$$H[f(x)] = \int_0^1 \log \frac{dF^{-1}(p)}{dp} dp, \quad (9)$$

considering the change of variables $x = F^{-1}(p)$. He proposed to estimate the cdf $F(X)$ with the empirical distribution function $F_n(x)$ and used the difference operator instead of the differential operator in order to approximate the $\frac{dF^{-1}(p)}{dp}$ by a function of the order statistics.

Suppose that X_1, \dots, X_n be a random sample from the distribution function $F(x)$ whose order statistics are $X_{(1)}, \dots, X_{(n)}$. Vasicek [10] proposed the following entropy estimator,

$$H_V = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (10)$$

with $m < n/2$ a positive integer number, $X_{(i+m)} - X_{(i-m)}$ the spacing of order m , or m -spacing, $X_{(i)} = X_{(i)}$ if $1 < i$ and $X_{(i)} = X_{(n)}$ if $i > n$. The author proves that H_V estimator is weakly consistency of $H(f)$ when $m/n \rightarrow 0$ and $n, m \rightarrow \infty$.

Several authors introduced modifications to the Vasicek estimator. In this work we consider, furthermore HV (10), the following entropy estimators which were presented in Al-Omari [13].

- Van Es [11]:

$$\begin{aligned} H_{VE} &= \frac{1}{n-m} \sum_{i=1}^{n-m} \log \left[\frac{n+1}{m} (X_{(i+m)} - X_{(i)}) \right] \\ &\quad + \sum_{k=m}^n \frac{1}{k} + \log \frac{m}{n+1}. \end{aligned} \quad (11)$$

- Correa [12]:

$$H_C = -\frac{1}{n} \sum_{i=1}^n \log \frac{\sum_{j=i-m}^{i+m} (j-i) (X_{(j)} - \bar{X}_{(i)})}{n \sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})^2}, \quad (12)$$

where $\bar{X}_{(i)} = (2m+1)^{-1} \sum_{j=i-m}^{i+m} X_{(j)}$.

- Noughabiet et al. [21]

$$H_{NA} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{c_i m} (X_{(i+m)} - X_{(i-m)}) \right] \quad (13)$$

where

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

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

- Al-Omari [22]:

$$H_{AO_1} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{\omega_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (14)$$

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

- Al-Omari alternative proposal [13]:

$$H_{AO_2} = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{v_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (15)$$

where

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

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

- Ebrahimi [15]

$$H_E = \frac{1}{n} \sum_{i=1}^n \log \left[\frac{n}{\tau_i m} (X_{(i+m)} - X_{(i-m)}) \right], \quad (16)$$

where

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

Van Es [11] showed that, under general conditions, this estimator converge almost surely to $H[f(x)]$ when $m, n \rightarrow \infty$, $m/\log(n) \rightarrow \infty$, and $m/n \rightarrow 0$, and also showed the asymptotic normality when $m, n \rightarrow \infty$ and $m = o(n^{1/2})$. Correa [12], through a simulated study, showed that his proposal has a smaller mean squared error than the Vasicek proposal (10). Al-Omari [22] presented three estimator based on different sampling method. In this paper we studied the simple

random sampling one (SRS) (14) and second proposal suggested in [13]. The author proved that these estimators converge in probability to $H[f(x)]$ when $m, n \rightarrow \infty$ and $m/n \rightarrow 0$. Ebrahimi et al [15] presented an estimator adjusting the Vasicek's [10] weight. Under the same conditions than Al-Omari [22] the authors proved that $HE \xrightarrow[n \rightarrow \infty]{p} H[f(x)]$ when $m, n \rightarrow \infty$ and $m/n \rightarrow 0$. The same applies to Noughabi et al. estimator [21].

V. SIMULATION STUDY

The choice of the spacing parameter m in this kind of estimators is an important task that is still open. Wiczorkowski et al. [23] proposed a heuristic formula $m = [\sqrt{n} + 0.5]$. To assess the performance of this criterion we conducted a Monte Carlo study for the entropy estimators presented in section IV-B under the \mathcal{G}^0 model.

We generated 1000 samples from a \mathcal{G}^0 distribution of size $n \in \{9, 25, 49, 81, 121\}$ for each textured value $\alpha \in \{-8, -5, -3, -1.5\}$ and $L = 2$. For each non-parametric study we use the m value chosen with the Wiczorkowski et al. proposal. The sample sizes chosen represents different scenarios of window sizes. Since the parameter γ is proportional to the brightness, meaning it is a scale parameter, we based the forthcoming analysis on the condition $E(Z) = 1$ which links texture and brightness parameter by $\gamma^* = -\alpha - 1$.

Each replication produces a vector of estimators $(\hat{H}_1, \dots, \hat{H}_{1000})$ from which we compute sample mean \bar{H} , sample bias $\hat{B} = B_{\hat{H}} = \bar{H} - H$ where H is the true entropy value, and sample mean squared error (MSE) $\widehat{\text{MSE}} = (1000)^{-1} \sum_{i=1}^{1000} (\hat{H}_i - H)^2$.

Figures 2(a) and 2(b) presented the bias and the MSE for the Wiczorkowski et al. [23] criterion, $L = 2$ case and for all of the estimators analyzed, except for Al-Omari (IV-B) and Ebrahimi (16) estimators due to the large bias they presented in the simulation study. It can be seen that there is no single estimator that performs well for all α value. H_C and H_{AO_1} present low bias and low MSE for all the cases study except for $\alpha = -1.5$. The others estimators show bad behavior for all the cases studied.

In order to improve the Wiczorkowski et al. [23] criterion we implement another strategy to choose, for each sample size n , the best m value to be used for all α values.

Table I shows the schema of the methodology employed, for fixed n value and entropy estimator. Each table entry, \hat{B}_{ij} , represents the bias for $m = i$ and $\alpha = \alpha_j$. For each estimator and each fixed m value we calculate the average bias (\bar{B}) between different α values. Finally, the chosen m is the argument that minimizes the average bias.

Table V shows the best m chosen according to the methodology used for $L = 1$ and $L = 2$ respectively, for samples coming from \mathcal{G}_0 distribution.

Figures 3(a) and 3(b) show the behavior of the estimators studied for the m value chosen in terms of bias and MSE respectively for $L = 2$ case. We also plot the H_{MV} estimator. It can be observed the improvement in entropy estimation in terms of bias and MSE with our methodology, compared with

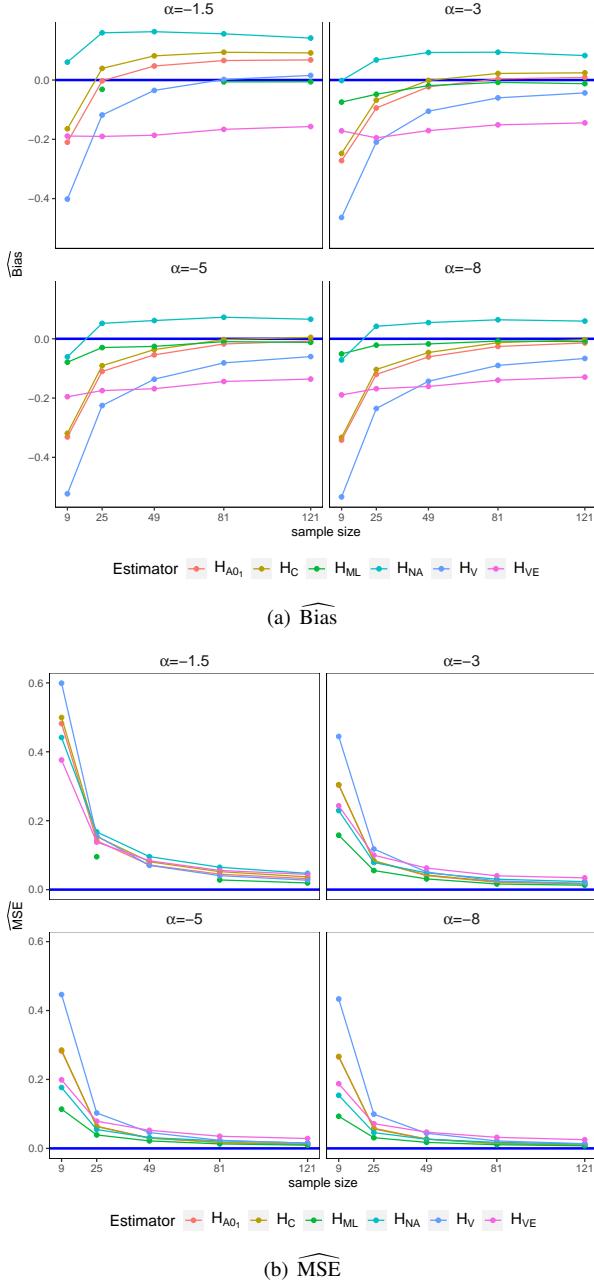


Fig. 2. Bias and MSE for Wieczorkowski criterion, $L = 2$.

the Wieczorkowski's et al. empirical formula for all of the estimators studied.

We continue studying the quality of these estimators in the SAR image classification framework.

VI. CLASSIFICATION

With the aim to study the performance of the selected entropy estimators in terms of SAR image classification, we divided the analysis into synthetic and actual images. We used unsupervised and supervised techniques to choose the three best estimators that present the best values of classification quality. For the former we carried out a k -means algorithm which groups data into k classes setting k centroids and minimizing the variance within each group. This nonhierarchical clustering technique is

TABLE I
SELECTION CRITERION FOR THE BEST m FOR EACH n AND EACH ENTROPY ESTIMATOR, WITH $\alpha_1 = -1.5$, $\alpha_2 = -3$, $\alpha_3 = -5$ AND $\alpha_4 = -8$.

m	α_1	α_2	α_3	α_4	$\bar{\hat{B}}$
1	\hat{B}_{11}	\hat{B}_{12}	\hat{B}_{13}	\hat{B}_{14}	$\bar{\hat{B}}_1$
2	\hat{B}_{21}	\hat{B}_{22}	\hat{B}_{23}	\hat{B}_{24}	$\bar{\hat{B}}_2$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	\hat{B}_{i1}	\hat{B}_{i2}	\hat{B}_{i3}	\hat{B}_{i4}	$\bar{\hat{B}}_i$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$\lfloor n/2 \rfloor$	\hat{B}_{N1}	\hat{B}_{n2}	\hat{B}_{n3}	\hat{B}_{n4}	$\bar{\hat{B}}_n$
$m = \arg \min_i \hat{B}_i$					

TABLE II
BEST m CHOSEN FOR EACH n AND ENTROPY ESTIMATOR.

L	n	H_{AO_1}	H_C	H_{NA}	H_V	H_{VE}
1	9	4	4	3	4	4
	25	6	5	3	8	2
	49	7	5	4	9	2
	81	7	4	5	8	2
	121	9	4	6	11	2
2	9	4	4	3	3	2
	25	8	4	4	9	2
	49	8	4	5	9	2
	81	9	4	5	9	2
	121	10	5	6	10	2

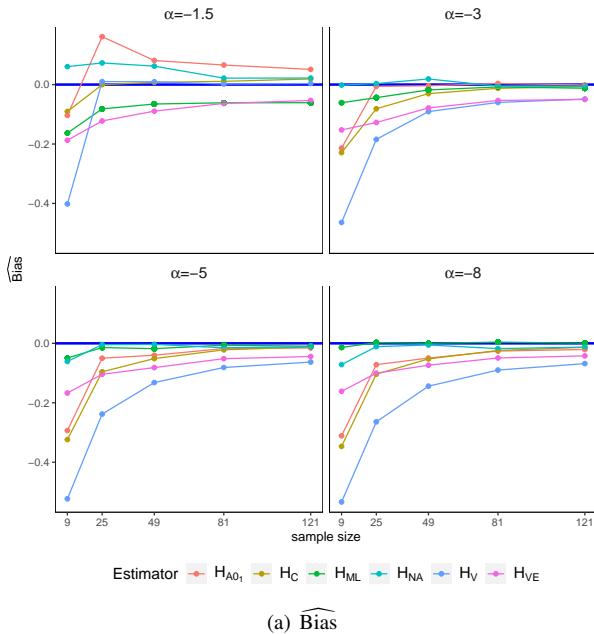
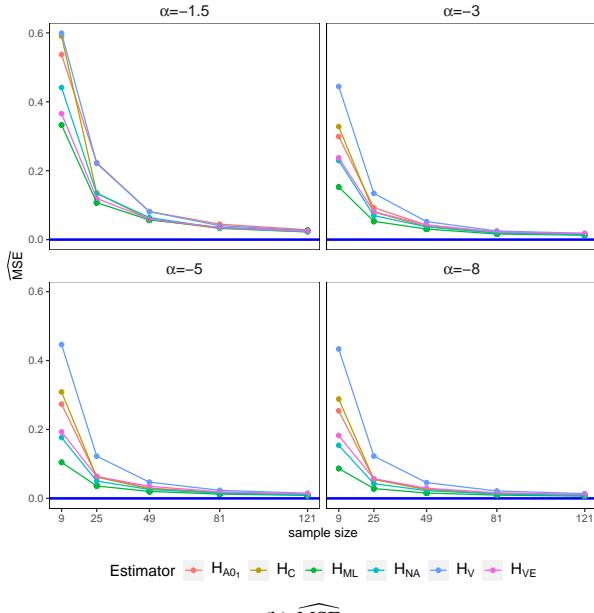
applied in many studies in SAR image processing. Niharika et al. [24] put forward a segmentation algorithm for SAR images based on k -means classification and thresholding techniques. Liu et al. [25] propose a k -means clustering technique for SAR image change detection.

For the latter approach we implemented a support vector machine (SVM) algorithm, which is a supervised machine learning technique based on the statistical learning theory [26] whose objective is to define, given a set of features, the best possible separation between classes by finding a hyperplane that maximizes the margin of separation between these classes. It is common to sacrifice some misclassification to get a better overall performance, introducing a penalizing parameter c . When data cannot be separated by a hyperplane, they are transformed to a higher dimensional feature space through a suitable nonlinear transformation called kernel function. Given $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$, linear and radial kernels are respectively defined by $K_L(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$ and $K_R(\mathbf{x}, \mathbf{x}') = \exp(-g\|\mathbf{x} - \mathbf{x}'\|^2)$, for $g > 0$.

This method has been widely applied in different areas such as sea oil spill monitoring [27], pattern recognition [28], classification of polarimetric SAR Image [9, 29] among other applications.

To measure the quality of these algorithms we use different measures depending on the type of classification. In the unsupervised case we use Calinski-Harabasz (CH) [30] and Davies-Bouldin [31] (DB) indexes, Kappa coefficients for the supervised classification and accuracy for both algorithms.

In the following, we present empirical results classifying a

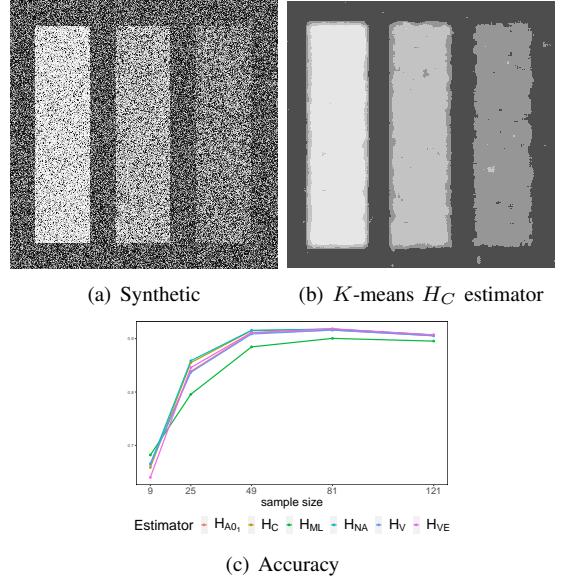
(a) $\widehat{\text{Bias}}$ (b) $\widehat{\text{MSE}}$ Fig. 3. Bias and MSE for authors proposal m choice, $L = 2$.

synthetic image SAR.

A. Synthetic image

Figure 4(a) shows a 300×300 image generated with a \mathcal{G}^0 distribution with $L = 2$, $\gamma = 0.1$ and four different classes according to the value of $\alpha \in \{-1.5, -3, -5, -8\}$. The light gray level corresponds to an extremely textured area corresponding to the value $\alpha = -1.5$. As the gray level decreases, the texture of the zone changes from heterogeneous ($\alpha = -3$ and -5) to a homogeneous zone corresponding to the darkest gray level ($\alpha = -8$).

For each estimator we perform a map of estimated entropies (\hat{H}) by sweeping the image with sliding windows of sizes $s \times s$, for $s = 3, 5, 7, 9, 11$, corresponding to the sample sizes

Fig. 4. K -means applied to a synthetic image with $L = 2$, $\gamma = 0.1$ and sliding windows size 9×9 .

studied in section V. \hat{H} is then used as a feature to classify in both, unsupervised and supervised technique.

For each windows size we applied k -means algorithm to the synthetic image. Figure 4(c) shows the accuracy according to the chosen sample size. It can be observed that a 9×9 window presents the best value of accuracy. It can also be seen that the H_{ML} estimator presents the worth performance whereas H_C , H_{NA} and H_{VE} show the best performance, as is indicated in Table III with bold letter. Figure 4(b) shows the result of applied a k -means classification for H_C estimator. Table IV presents the value of CH and DB indexes for $n = 81$ sample size. It can be observe that for the former case, H_C , H_{NA} and H_{VE} estimators have the best performance, and for the latter H_C , H_{NA} and H_V were the best.

TABLE III
ACCURACY FOR k -MEANS ($k = 4$) APPLIED TO SIMULATED DATA

n	H_{AO_1}	H_C	H_{NA}	H_V	H_{VE}	H_{ML}
9	0.664	0.659	0.665	0.665	0.640	0.682
25	0.839	0.855	0.859	0.837	0.846	0.796
49	0.911	0.915	0.915	0.909	0.911	0.884
81	0.916	0.918	0.918	0.916	0.918	0.900
121	0.905	0.906	0.907	0.905	0.907	0.895

TABLE IV
CLASSIFICATION QUALITY INDEXES FOR k -MEANS ($k = 4$) APPLIED TO SIMULATED DATA WITH $n = 81$

Index	H_{AO_1}	H_C	H_{NA}	H_V	H_{VE}	H_{ML}
CH	852914	898079	902719	852914	867746	703774
DB	0.441	0.434	0.433	0.441	0.442	0.467

We also applied supported vector machine (SVM) algorithm using each entropy estimator as the contributor feature, to the synthetic image under the same conditions as for the k -means algorithm. We also analyzed the H_{ML} estimator.

To find the best kernel and hyperparameters we took at random 1000 pixels in each of the four regions, far away enough from the boundaries. This reference sample was divided into two sets: training and validation (80 % of the sample), and testing (20 %). We considered linear and radial types for the kernel, cost of restrain violation $c = 0.001, 0.01, 0.1, 1, 5, 10$ and $g = 0.01, 0.1, 1, 1.5, 2$. With the 80 % of the training-validation set we made k -cross fold validation, with $k = 5$, and computed the mean and the standard deviation of the F1-scores. Recall that for True Positive Rate (TPR) and Positive Predictive Value (PPV), it is defined $F1 = 2 \cdot TPR \cdot PPV / (TPR + PPV)$. Table V shows the selected kernels and hyperparameters that maximize F1 mean and minimize F1 variance. In presence of matching, we decided to opt for the fastest or simplest choice.

Once we found the best models, we computed the accuracy and the κ coefficient using the remaining 20 % of the reference samples. The results are exhibited in Figure 5(a). Finally, models were trained using the whole reference sample and applied to classify the complete image. The same measures were computed and the results are shown in Figure 5(b).

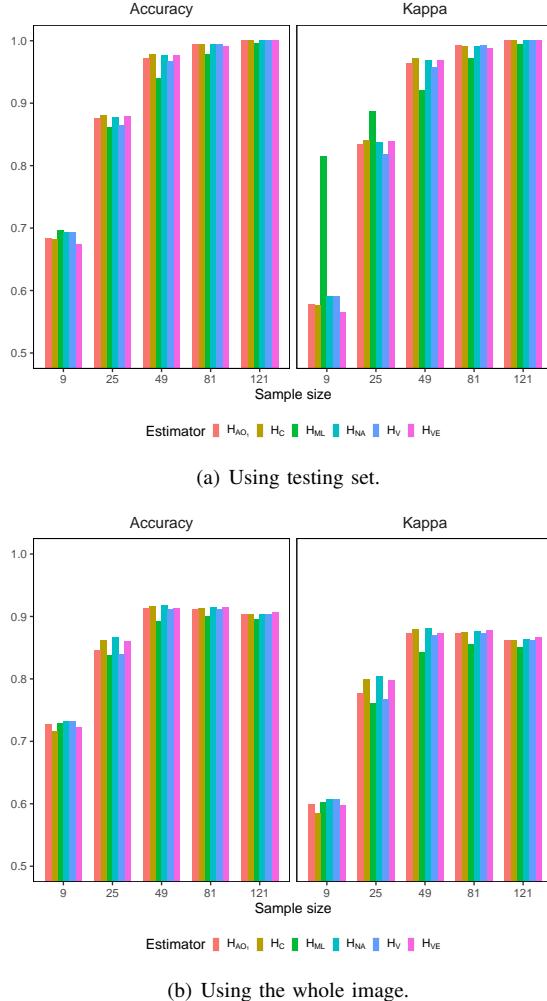


Fig. 5. Accuracy and Kappa coefficient for SVM.

It can be seen that the best performance is achieved for sliding windows of size 9×9 and H_C , H_{NA} and H_{VE} estimators with 0.914, 0.914 and 0.915 respectively for accuracy

values, and 0.876, 0.877 and 0.878 for Kappa index, when the classifier is applied to the whole image.

Since these estimators and window size are the ones that performed the best in the synthetic image, we apply them in the classification of an actual image.

B. Actual image

We used a subsample of 500×645 pixels from a fully PolSAR image of 900×1024 data of the San Francisco bay area in California, taken by the NASA/JPL AIRSAR L-band instrument, in intensity format where the number of looks $L = 4$. This can be appreciated in Figure 6.

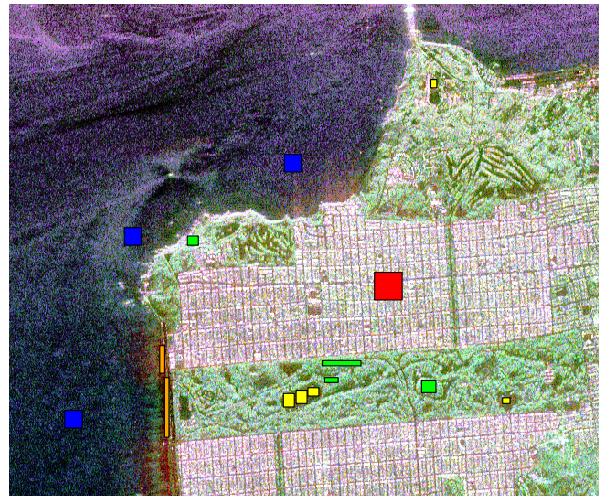


Fig. 6. Image of San Francisco with reference samples.

The equivalent number of looks (ENL) using uncorrelated data is defined as $ENL = 1/\widehat{CV}^2$, the reciprocal of the sample coefficient of variation $\widehat{CV} = \widehat{\sigma}/\widehat{\mu}$, where $\widehat{\sigma}$ is the sample standard deviation and $\widehat{\mu}$ the sample mean [32]. In order to find the ENL in each polarization band, we manually selected samples from homogeneous areas in each band and calculated ENL as an average weighted by the sample size per band. Finally, ENL is the average of the estimations in each polarization. We obtained 2.53, 3.41 and 3.41 as the ENL values in HH, HV and VV band respectively. For these reasons we used the m values for $L = 2$ which are the same than $L = 3$.

We applied a SVM algorithm to the image of San Francisco replicating the procedure described in the study of simulated data, using the entropy estimator as a feature for classification the three polarizations.

Figure 6 shows the training samples selected to perform the supervised classification. We selected five areas which consist of water (blue), urban zone (red), vegetation (green), pasture (yellow) and beach (orange).

We studied lineal and radial kernels, but the latter produced the best results with the following hyperparameters:

- $c = 10$ and $g = 1.5$ for H_C ,
- $c = 5$ and $g = 2$ for H_{NA} ,
- $c = 10$ and $g = 2$ for H_{VE} .

TABLE V
BEST KERNEL (L: LINEAL, R:RADIAL) AND HYPERPARAMETERS FOR SVM APPLIED TO SYNTHETIC SAR DATA.

n	H_{AO_1}	H_C	H_{NA}	H_V	H_{VE}	H_{ML}
9	R, $c = 5, g = 1$	R, $c = 1, g = 0.1$	R, $c = 5, g = 2$	R, $c = 5, g = 2$	L, $c = 0.1$	R, $c = 10, g = 0.01$
25	R, $c = 5, g = 2$	L, $c = 10$	L, $c = 10$	R, $c = 10, g = 1$	R, $c = 1, g = 1.5$	L, $c = 0.1$
49	L, $c = 10$	L, $c = 10$	R, $c = 10, g = 1.5$	R, $c = 5, g = 1.5$	L, $c = 0.1$	R, $c = 5, g = 1$
81	R, $c = 10, g = 2$	L, $c = 5$	L, $c = 10$	R, $c = 10, g = 2$	R, $c = 1, g = 2$	L, $c = 1$
121	L, $c = 5$	L, $c = 5$	L, $c = 1$	L, $c = 1$	L, $c = 1$	L, $c = 0.01$

Moreover, in a second stage we included the CV as a feature in the classification process, which showed an improvement in the quality measures. In this case, the best performance is achieved for linear kernel with cost equals to 10.

Table VI presents the accuracy and Kappa index. We can observe that the classification is improved when we consider CV coefficient. In terms of entropy estimators, H_{NA} seems to be the best classifier including the CV as a feature.

TABLE VI
ACCURACY AND KAPPA COEFFICIENT VALUES IN THE TESTING SET FOR THE IMAGE OF SAN FRANCISCO.

Training with entropy estimators		Training with entropy estimators and cv	
Accuracy	Kappa	Accuracy	Kappa
H_C	0.9681	0.9596	0.9917
H_{NA}	0.9669	0.9581	0.9929
H_{VE}	0.9705	0.9626	0.9906

Figure 7 exhibits the classification of the whole image of San Francisco when our proposal is applied. It can be observed that the classifier distinguishes the beach and, with the addition of the CV, some roads surrounded by trees are better classified.

VII. CONCLUSIONS

We assess the performance of six non parametric entropy estimators in conjunction with ML estimator, not only in terms of bias and MSE but also in terms of image classification.

The advantage of using these non-parametric estimators is that they are very simple to implement, they do not assume any model and they do not need to use any optimization algorithm. The disadvantage is that they depend on a spaced parameter m . We solved these point by defining a criterion to choose the best m value that present the least bias in the entropy estimation for all the textured values studied and all sample sizes analyzed. This criterion presents a better performance than that of Wieczorkowski et al. [23].

With these m values we implemented an unsupervised (k -means) and supervised (SVM) classification algorithm in synthetic and actual images and compared their performance with the ML entropy estimator H_{ML} . We displayed that H_C , H_{NA} and H_{VE} show the best performance on most of the classification measures used.

For these estimators we study their performance applying a supervised classification to a real image. We concluded that H_{VE} presents the best accuracy an Kappa values, 97 % and 96 % respectively.

In addition, we also considered a new feature, the CV coefficient and found that both, accuracy and Kappa index showed

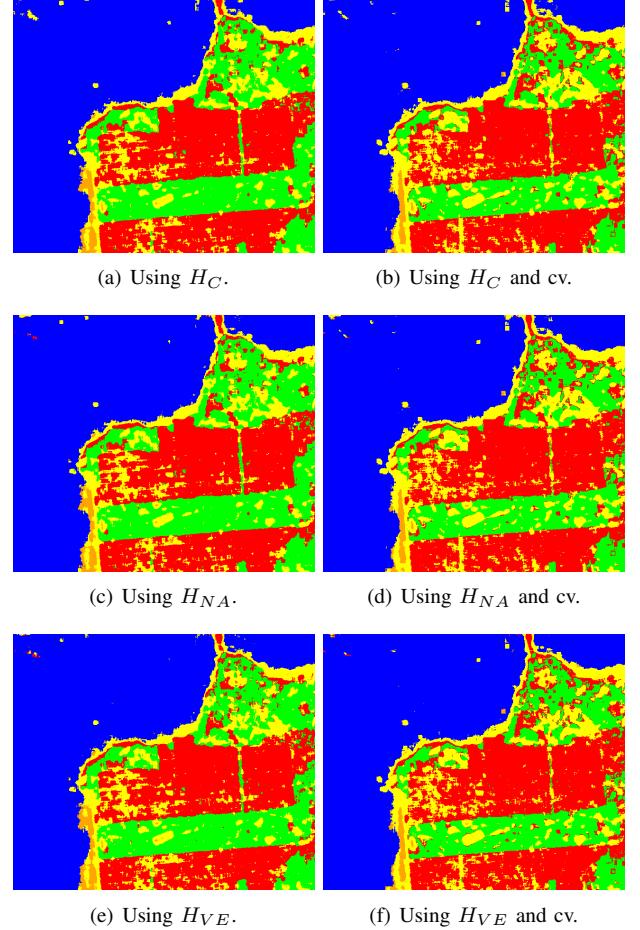


Fig. 7. San Francisco classification.

an improvement of 2 % in the classification performance and that the H_{NA} estimator presented the best behaviour for the studied measures.

VIII. COMPUTATIONAL INFORMATION

All studies were made in the R platform and language for statistical computing [33] (version 4.0.4).

REFERENCES

- [1] A. C. Frery, H. J. Müller, C. C. Yanasse, and S. J. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 648–659, 1997.
- [2] J. Gambini, J. Cassetti, M. Lucini, and A. Frery, "Parameter estimation in SAR imagery using stochastic distances and asymmetric kernel," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, pp. 365–375, 2015.

- [3] N. Carvalho, L. Sant'Anna Bins, and S. J. Sant'Anna, "Analysis of stochastic distances and Wishart mixture models applied on PolSAR images," *Remote Sensing*, vol. 11, no. 24, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/24/2994>
- [4] A. C. Frery, R. J. Cintra, and A. D. Nascimento, "Entropy-based statistical analysis of PolSAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 6, pp. 3733–3743, 2013.
- [5] A. D. Nascimento, R. J. Cintra, and A. C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 373–385, 2010.
- [6] A. D. Nascimento, M. M. Horta, A. C. Frery, and R. J. Cintra, "Comparing edge detection methods based on stochastic entropies and distances for PolSAR imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 2, pp. 648–663, 2014.
- [7] R. H. Nobre, F. A. A. Rodrigues, R. C. P. Marques, J. S. Nobre, J. F. S. R. Neto, and F. N. S. Medeiros, "SAR image segmentation with Rényi's entropy," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1551–1555, 2016.
- [8] J. Ferreira and A. D. Nascimento, "Shannon entropy for the \mathcal{G}_I^0 model: A new segmentation approach," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2547–2553, 2020.
- [9] M. G. Palacio, S. B. Ferrero, and A. C. Frery, "Revisiting the effect of spatial resolution on information content based on classification results," *International Journal of Remote Sensing*, vol. 40, no. 12, pp. 4489–4505, 2019. [Online]. Available: <https://doi.org/10.1080/01431161.2019.1569278>
- [10] O. Vasicek, "A test for normality based on sample entropy," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 38, no. 1, pp. 54–59, 1976. [Online]. Available: <http://www.jstor.org/stable/2984828>
- [11] B. van Es, "Estimating functionals related to a density by a class of statistics based on spacings," *Scandinavian Journal of Statistics*, vol. 19, no. 1, pp. 61–72, 1992. [Online]. Available: <http://www.jstor.org/stable/4616225>
- [12] J. C. Correa, "A new estimator of entropy," *Communications in Statistics – Theory and Methods*, vol. 24, no. 10, pp. 2439–2449, 1995. [Online]. Available: <https://doi.org/10.1080/03610929508831626>
- [13] A. I. Al-Omari, "A new measure of entropy of continuous random variable," *Journal of Statistical Theory and Practice*, vol. 10, no. 4, pp. 721–735, 2016. [Online]. Available: <https://doi.org/10.1080/15598608.2016.1217444>
- [14] H. A. Noughabi and R. A. Noughabi, "On the entropy estimators," *Journal of Statistical Computation and Simulation*, vol. 83, no. 4, pp. 784–792, 2013. [Online]. Available: <https://doi.org/10.1080/00949655.2011.637039>
- [15] N. Ebrahimi, K. Pflughoefl, and E. S. Soofi, "Two measures of sample entropy," *Statistics & Probability Letters*, vol. 20, no. 3, pp. 225–234, 1994. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0167715294900469>
- [16] M. Mejail, J. C. Jacobo-Berelles, A. C. Frery, and O. H. Bustos, "Classification of SAR images using a general and tractable multiplicative model," *International Journal of Remote Sensing*, vol. 24, no. 18, pp. 3565–3582, 2003.
- [17] M. E. Mejail, A. C. Frery, J. Jacobo-Berelles, and O. H. Bustos, "Approximation of distributions for SAR images: Proposal, evaluation and practical consequences," *Latin American Applied Research*, vol. 31, pp. 83–92, 2001.
- [18] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [19] D. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Springer Science+Business Media, LLC, 2008.
- [20] G. Casella and R. Berger, *Statistical Inference*. Duxbury Resource Center, June 2001.
- [21] A. H. Noughabi and N. R. Arghami, "A new estimator of entropy," *Journal of the Iranian Statistical Society*, vol. 9, no. 1, 2010. [Online]. Available: <http://jirss.irstat.ir/article-1-81-en.html>
- [22] A. I. Al-Omari, "Estimation of entropy using random sampling," *Journal of Computational and Applied Mathematics*, vol. 261, pp. 95–102, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0377042713006006>
- [23] R. Wieczorkowski and P. Grzegorzewski, "Entropy estimators-improvements and comparisons," *Communications in Statistics - Simulation and Computation*, vol. 28, no. 2, pp. 541–567, 1999. [Online]. Available: <https://doi.org/10.1080/03610919908813564>
- [24] E. Niharika, H. Adeeba, A. S. R. Krishna, and P. Yugander, "K-means based noisy SAR image segmentation using median filtering and Otsu method," in *2017 International Conference on IoT and Application (ICIOT)*, 2017, pp. 1–4.
- [25] L. Liu, Z. Jia, J. Yang, and N. Kasabov, "Sar image change detection based on mathematical morphology and the k-means clustering algorithm," *IEEE Access*, vol. 7, pp. 43 970–43 978, 2019.
- [26] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995.
- [27] J. Fan, F. Zhang, Z. Dongzhi, and J. Wang, "Oil spill monitoring based on SAR remote sensing imagery," *Aquatic Procedia*, vol. 3, 03 2015.
- [28] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [29] L. Zhang, Z. Bin, J. Zhang, and Y. Zhang, "Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, 01 2010.
- [30] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/03610927408827101>
- [31] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, 1979.
- [32] S. Anfinsen, A. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3795–3809, 2009.
- [33] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: <https://www.R-project.org/>