Enhancing Change Detection in Ultra-Wideband VHF SAR Imagery: An Entropy-Based Approach With Median Ground Scene Masking

João Gabriel Vinholi[ID], Paulo Ricardo Branco da Silva[ID], Dimas Irion Alves[ID], *Member, IEEE*, and Renato Machado[ID], *Senior Member, IEEE*

Abstract—We propose an algorithm based on information theory to detect changes in ultra-wideband (UWB) very-high frequency (VHF) synthetic aperture radar (SAR) images with high performance and low complexity. Our algorithm models the clutter-plus-noise using six different distributions and computes a scalar statistic for each pixel based on a multitemporal stack of images. With this statistic, it is then possible to apply hypothesis testing and classification methods to infer the occurrence of a change. In this context, we derive expressions necessary for the entropy-based statistics, including the entropy variance for the Weibull and Rice distributions. We also evaluate the computational time complexity of the algorithm for each distribution studied. Furthermore, a masking strategy is used to reduce false alarms significantly. We show that the mask mapping assumptions are mild in scenarios with stacks of images, allowing its use in many scenarios. Our algorithm achieves a false alarm rate (FAR) of 0.08 and a probability of detection (PD) of 100%, outperforming existing methods on the CARABAS II dataset.

Index Terms—Change detection, entropy, information theory, stochastic distances, synthetic aperture radar (SAR).

I. Introduction

CHANGE detection is a remote sensing task that involves detecting changes in an area of interest using two or more multitemporal images. This allows the detection of appearance changes, object disappearance or appearance, or even changes in the scene's background [1]. Change detection has many applications, such as monitoring land use, urban growth, deforestation, disaster recovery, and military surveillance [2]. This article addresses the problem of detecting hidden targets in forested areas using ultra-wideband (UWB) very-high frequency (VHF) wavelength-resolution synthetic aperture radar (SAR) images. UWB VHF sensors have the advantage of penetrating dense foliage and capturing large and stable scatterers. We propose a change detection method that compares images taken at different times and identifies changes related to new targets concealed under the vegetation. Our method ignores other changes that are irrelevant or caused by acquisition or environmental factors.

Images obtained by SAR tend to be favored over optical images for foliage penetrating (FOPEN) applications because of good resolution for human-made objects, lower susceptibility to weather conditions, and the mitigation of undesirable vegetation aspects [3], [4], favoring the detection of changes otherwise invisible.

There has been wide use of SAR images in applications such as marine analysis, agriculture, and forestry. In [5], high-spatial-resolution flow measurements are derived for the Kuroshio Current in the East China Sea from SAR data using a theoretical model of shifts in Doppler frequency. In [6], a time series of vertical–horizontal polarized Sentinel-1A SAR scenes is processed. A decision tree algorithm is used to analyze topographical features from the SAR images in conjunction with rice phenological parameters, such as the transplanting date, mature grain date, length of growing season, and the green-up speed, in order to map patches of rice crops. In [7], a multilevel model for forest height inversion is introduced and investigated using X-band single-pass interferometric SAR coherence data.

The problem with these applications, though, is that the more standard microwave frequencies used in SAR imagery result in many false alarms when aimed at large targets [8]. Specifically in the forestry domain, such frequency ranges seem to produce poor results for biomass estimation, because reflectance in these frequencies only partially correlates to woody mass [9]. In [9], it is stated that at higher frequencies, such as at the X- and C-bands, the high attenuation of the canopy leads to scattering being largely composed of the effects of small structures in the crown of the trees. This makes the saturation levels of the backscatter be attained at

very low levels of biomass. Lower frequencies, such as the L- and P-bands, show improved sensitivity, since the larger wavelengths contribute to the scattering from trunks and large branches, which contain most of the effective woody biomass. However, even at these bands, the saturation effect happens for relatively low quantities of biomass [10]. Findings like these suggest that lower frequencies are desirable for the penetration of dense canopies. Results from two aerial VHF-band SAR measurement campaigns, known as the CARABAS I and II [11], indicate that the VHF-band backscatter is more sensitive to biomass detection than the P-band as a result of its higher saturation threshold to biomass [12], [13].

As hinted above, UWB VHF SAR systems, e.g., the CARABAS II system [11], can be used to mitigate the issues of higher frequency ranges [8]. They have a large fractional bandwidth and a wide antenna bandwidth, yielding system resolutions in the order of the radar signal wavelengths [11], hence the denomination "wavelength resolution." In addition, as it was observed for the small structures in the crowns of trees, these systems are not very sensitive to small scatterers in the ground scene, i.e., to objects with dimensions smaller than the signal wavelength [9], [14]. A consequence is that images suffer little from speckle noise, since there might be only a single scatterer in the resolution cell [15]. Moreover, large scatterers tend to be static objects that are not very sensitive to weather conditions. This allows acquiring similar images for a determined ground area via multiple image passes [15]. This agrees with the observations made in [9] and points to VHF wavelength-resolution SAR images being adequate for FOPEN applications [16], specifically for the detection of stable and relatively large concealed targets in forested areas using change detection methods [17], [18], [19], [20], [21], [22], [23].

A common approach in change detection methods for VHF wavelength-resolution SAR images is the pairwise approach, where statistical models are employed to represent the clutter-plus-noise distribution in surveillance and reference images. These models are usually coupled with the Neyman–Pearson (NP) criterion in the creation of decision functions [24]. This is the approach of [17], which simplifies hypothesis tests to space–time equations. Similarly, [25] uses statistical hypothesis tests based on Izawa's bivariate gamma distribution. Change detection can also be performed using difference images. In [26], these images are processed with adaptive noise cancelers. Bayes' theorem can also be used to estimate the probability of change based on target or clutter-plus-noise probability distributions using histograms for the conditional reference and surveillance clutter-plus-noise probabilities [20], [22], [23].

Statistical analysis of wavelength-resolution SAR image stacks can reduce the number of false alarms concerning the traditional pairwise approach [19]. In [21], the NP criterion is used on statistical models for the background clutter and the targets in an image stack. A stack-based method based on robust principal component analysis has been shown to produce detection with low false alarm rates (FARs) [27]. In addition, the use of convolutional neural networks on image stacks achieves considerable reductions in the number of false

alarms [28], [29], [30]. While these methods have successfully reduced false alarms, our article seeks to improve performance by investigating a novel change detection methodology.

Thus, the main objective of this article is to use a change detection methodology not yet explored for stacks of wavelength-resolution SAR images and employ it to reduce the FARs presented in articles such as [17], [19], [20], [21], and [25]. In particular, we use an algorithm based on the information-theoretic quantity known as the differential Shannon entropy. This quantity is a special case of the (h, ϕ)-entropy, defined in [31]. In [32], under the assumption of MLE of distribution parameters, entropy becomes a random variable (RV) where its probability density function (pdf) asymptotically converges to a Gaussian distribution. This allows the use of hypothesis tests over entropy-based statistics. This approach has been used for change detection in [33], where hypothesis testing is conducted on polarimetric SAR (POLSAR) data. In [33], the test statistics are squared differences of entropy measurements for Wishart-modeled POLSAR data.

In this article, we propose a novel change detection methodology for wavelength-resolution SAR images based on information-theoretic measures. Our methodology consists of four steps: 1) computing a stack entropy statistic for each pixel using different distributions that can model the clutter-plus-noise; 2) applying a median-based hypothesis test to detect changes; 3) performing morphological operations to refine the change map; and 4) evaluating the performance using false alarm probability p_{FA} and detection probability p_{D}. We also derive expressions for the computational complexity of our algorithm.

For the asymptotic case in which the number of samples tends to infinity, we obtain the variance of the entropy for distributions that had not been previously considered in the Literature of entropy-based change detection, namely the Weibull and Rice distributions. Inspired by [18]—in which the stackwise median operation is used to generate estimations of the ground scene of UWB VHF SAR images—we also use the median statistic as a tool to reduce FAR. In particular, the median of the entropies of the images contained in a stack is used as a ground scene mask, which efficiently masks out false alarms. Our experimental results demonstrate that the proposed algorithm outperforms existing methods in terms of FAR, making it a valuable addition to the change detection literature.

In what follows, we list our main contributions.

1) A novel change detection algorithm (CDA) is presented for UWB VHF images, which uses MLE to fit the data to particular pdfs and the subsequent use of the fit distributions in the calculation of entropy measures. These measures are turned into statistics for the changes in scenes. A comprehensive comparison shows that the proposed algorithm compares well with state-of-the-art methods [18], [19], [20], [21], performing better or at least equivalently to all of them.

2) A time-complexity analysis of the proposed algorithm demonstrates its efficiency and scalability with regard to the number of images, stacks, and pixels within a sliding window. The analysis shows that certain distributions,

such as the Gaussian and Rayleigh, are the most computationally efficient.

3) A masking scheme based on the median with regard to the stack dimension is used to reduce false alarms, increasing the recall of the original method.

The remainder of this article is divided as follows. In Section III, the use of entropy for hypothesis testing is discussed. In Section IV, pdfs that can model the statistics of the clutter and noise are discussed, and entropy variances are obtained for them. In Section V, we discuss our entropy-based CDA, splitting it into three parts: entropy-based statistic computation, median-based hypothesis testing, and morphological operations. In Section VI, simulations for stacks defined by flight pass numbers (stacks of size $M = 4$) and stacks defined by flight geometry angle (stacks of size $M = 8$) are analyzed, and results are discussed. Section VII concludes this article by discussing the key findings and their implications.

Notation: Lower and upper boldcase symbols represent vectors and matrices, respectively.

II. OVERVIEW OF THE CARABAS II DATASET

The CARABAS II is a UWB VHF SAR system for foliage penetration and target detection [34]. It transmits HH-polarized radio waves in 20–90 MHz, with wavelengths from 3.3 to 15 m. The resolution cell in the slant range is 2.5×2.5 m, suitable for target recognition. Most radar backscatter at the low VHF band comes from large scatterers, such as tree trunks, boulders, houses, and vehicles [34]. These objects have stable radar signatures between different image acquisitions.

As described in [34], CARABAS II is composed of 24 incoherent, calibrated images in the slant geometry. Each image contains 25 military vehicles camouflaged by vegetation. The images are subdivided into four distinct target deployments called missions, as indicated in Fig. 1. Each deployment was collected from six different passes, i.e., airborne SAR acquisitions, with three distinct flight geometry angles. Given that the northward direction is defined as angle 0° and angles increase counterclockwise, the flight geometry angles are 135°, 225°, and 230° [34]. There are two passes for each of the three flight geometry angles, resulting in six passes for each mission. This variety of flight passes and geometry angles allows a thorough evaluation of the algorithm's performance under different conditions. Its four missions are labeled with indices in the set {2, 3, 4, 5}. For consistency with the CARABAS II literature, we will adhere to this convention.

III. ENTROPY AS A STOCHASTIC MEASURE OF DISTANCE

In this section, we discuss the use of entropy in detecting changes among SAR magnitude images. We focus on describing a statistic obtained from entropy, with which it is possible to conduct hypothesis testing.

Let X be an RV with a probability distribution given by $f_X(x \in \mathcal{X}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$ is the length-p vector of deterministic distribution parameters, $\Theta \subseteq \mathbb{R}^p$ is the space of all possible length-p parameter vectors, and $\mathcal{X}$ is the support of X. Let either $\phi : [0, \infty) \to \mathbb{R}$ be a concave function and $h : \mathbb{R} \to \mathbb{R}$ be an ascending function or $\phi : (-\infty, 0] \to \mathbb{R}$ be

a convex function and $h : \mathbb{R} \to \mathbb{R}$ be a descending function. In [32], Salicrú et al. developed a general framework for entropy with (h, ϕ)-entropy classes. Under this framework, a general definition of differential entropy is given by

$$H_\phi^h(\boldsymbol{\theta}) = h\left(\int_{\mathcal{X}} \phi(f_X(x; \boldsymbol{\theta}) dx)\right). \tag{1}$$

In particular, the Shannon differential entropy of X is given by (1) when $h(z) = z$, $\phi(z) = -z \ln(z)$ and $\mathcal{X}$ is the set of real numbers, resulting in

$$H(\boldsymbol{\theta}) = -\int_{-\infty}^{\infty} f_X(x; \boldsymbol{\theta}) \ln(f_X(x; \boldsymbol{\theta})) dx. \tag{2}$$

Let $\hat{\boldsymbol{\theta}}$ be the MLE estimate of $\boldsymbol{\theta}$, that is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln(f_X(x; \boldsymbol{\theta})).$$

From [32], the difference in entropy $H(\hat{\boldsymbol{\theta}}) - H(\boldsymbol{\theta})$ converges in distribution to a Gaussian distribution. Specifically

$$\frac{1}{\sqrt{N}}\left(H(\hat{\boldsymbol{\theta}}) - H(\boldsymbol{\theta})\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_H^2\right) \tag{3}$$

where $\sigma_H^2 = \boldsymbol{\delta}^{\mathrm{T}} \mathbf{I}(\boldsymbol{\theta})^{-1} \boldsymbol{\delta}$ is the variance of the entropy. The vector $\boldsymbol{\delta} = \partial H(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is the gradient of the entropy with respect to vector parameter $\boldsymbol{\theta}$ and $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix, defined as

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}_X\left[\frac{\partial^2 \ln f_X(x; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right].$$

Let there be n different images with corresponding entropy measurements. For simplicity, in what follows, we will omit the dependence of H on $\boldsymbol{\theta}$. In addition, the entropy of each of the n images will be denoted by its respective integer. The null and alternative hypotheses are

$$\mathcal{H}_0 : H_1 = H_2 = \cdots = H_n \tag{4}$$

$$\mathcal{H}_1 : H_i \neq H_j \text{ for some } i \neq j, \; i, j \in \{1, \ldots, n\}. \tag{5}$$

From (3), assuming $\mathcal{H}_0$ and defining $H_1 = H_2 = \cdots = H_n = H$, [32] concludes that

$$\sum_{i=1}^n N_i\left(\frac{H_i(\hat{\boldsymbol{\theta}}_i) - H}{\sigma_{H_i}}\right)^2 \xrightarrow{\mathcal{D}} \chi_n^2 \tag{6}$$

where the convergence in (6) is in distribution and N_i is the number of samples for the ith image,[1] where $i \in \{1, 2, \ldots, n\}$.

However, the value of H cannot be observed, as it is a population rather than a sample statistic. Hence, it must be estimated. In [32], the sample mean

$$\overline{H} = \frac{1}{n}\sum_{i=1}^n H_i(\hat{\boldsymbol{\theta}}_i) \tag{7}$$

is used as the estimator of H.

---

[1]The computation of entropy and, potentially, of masking, thresholding, and morphological operations transforms the pixels in the original images into scalar statistic elements in 2-D grids. Even though it is not accurate to refer to the results of our computations as pixels and images, we will occasionally use this terminology, particularly when discussing morphological operations, because of the close relation our grid elements bear to pixels and for the sake of clarity and conciseness.
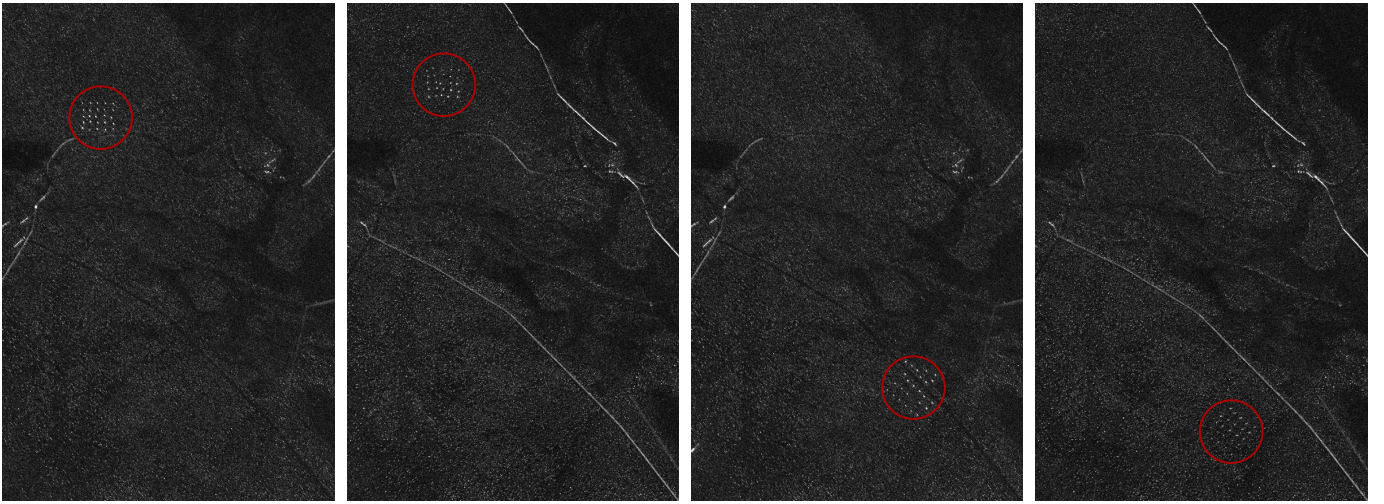
Fig. 1. Sample of the CARABAS II dataset. The four images here are from the same flight direction. They show the 25 target vehicles (in red) in the four possible deployments.

The use of $\overline{H}$ in place of H changes the probability distribution to which the LHS of (6) converges. According to [31]

$$\sum_{i=1}^{n} N_i \frac{\left(H_i\left(\hat{\boldsymbol{\theta}}_i\right) - \overline{H}\right)^2}{\sigma_{H_i}^2} \xrightarrow{\mathcal{D}} \chi_{n-1}^2. \tag{8}$$

For reference in the algorithmic description of change detection in Section V, we define the statistic e as

$$e \triangleq \sum_{i=1}^{n} N_i \frac{\left(H_i\left(\hat{\boldsymbol{\theta}}_i\right) - \overline{H}\right)^2}{\sigma_{H_i}^2}. \tag{9}$$

Note: Since entropy calculations require only the spatial dimensions and are performed on individual images, the probability distributions analyzed in Section IV for the description of clutter and noise correspond only to single images. This contrasts with typical stack analysis in the Literature (see [19] and [21, Chapters 4 and 5]), which considers statistical inference not only on the spatial (azimuth and range) axes of the image but also on the temporal axis.

IV. PDF ANALYSIS

In this section, pdfs that can describe the statistical properties of the radar intensity profiles of UWB VHF SAR images are briefly explained. Based on considerations such as the physics of UWB VHF SAR image generation, statistical hypothesis testing, and computational tractability, a series of pdfs are tested for the entropy description of the clutter-plus-noise.

Flexible distributions with more than two parameters, such as the generalized Gamma distribution (GΓD) [35] and the $\mathcal{G}_{AO}$ [36], are efficient in describing SAR images in the L, S, C, and X bands. In [35], the GΓD distribution is used to create a constant false alarm rate (CFAR) detection algorithm for high-resolution SAR images. Its efficiency is verified on SAR images in the X-band obtained from the TerraSAR-X satellite. In [36], the $\mathcal{G}_{AO}$ distribution addresses a situation not well described by the GΓD distribution, namely,

when the third-order sample log-cumulant of SAR data is close to zero. This makes the $\mathcal{G}_{AO}$ distribution a good fit for extremely heterogeneous land clutter, such as the one observed in urban regions. Experiments of the $\mathcal{G}_{AO}$ distribution on C-band RADARSAT-2 and L-band ALOS-PALSAR SAR data demonstrate that this distribution is effective for these bands [36].

However, distributions suitable for describing SAR images in super-UHF frequency ranges are not known to describe the statistical behavior of UWB VHF SAR images in the bidimensional range × azimuth domain. For that reason, instead of choosing distributions such as the GΓD or the $\mathcal{G}_{AO}$, we have used the distributions that have been shown to be good statistical fits in UWB VHF SAR image analysis.

According to [15], the clutter-plus-noise in images obtained from the difference between two wavelength-resolution amplitude SAR images with similar flight angles can be modeled by a Gaussian distribution. Since the Gaussian distribution is also tractable regarding the derivation of information-theoretical quantities, it is a reasonable choice for the pdf of the clutter-plus-noise and the implementation of entropy-based change detection.

The work in [37] shows via Anderson Darling goodness-of-fit tests that the Gamma distribution is a good candidate distribution for the modeling of clutter-plus-noise in wavelength-resolution intensity SAR difference images. In [25], it is shown that the bivariate Gamma distribution is a good model for the magnitudes of high-resolution, wavelength-resolution SAR images. In addition, [38] investigates the stability of change detection results obtained with statistical hypothesis tests for wavelength-resolution SAR change detection using the bivariate Rayleigh distribution. The bivariate Rayleigh distribution is also analyzed in [39], where it is compared to the bivariate K distribution. The work in [40] presents an iterative change detection method based on Bayes' theorem for UWB VHF SAR images considering commonly used clutter-plus-noise statistical models. The bivariate Rayleigh and bivariate Gaussian distributions

are investigated as candidates to model the clutter-plus-noise. Different aspects related to the distributions are discussed, and the impact of the distribution chosen for the proposed iterative change detection method is analyzed. Given the extensive amount of research on the Gaussian, Rayleigh, and Gamma distributions regarding the statistics of wavelength-resolution SAR images, we have opted to study these distributions in this article. The close connections between the log-normal and the Gaussian distributions, the Rice and the Rayleigh distributions, and the Weibull and the Rayleigh and exponential distributions inspired us also to choose to explore these distributions.

Another analysis on which we took inspiration was the work in [21], which uses the Anderson–Darling test to determine the goodness-of-fit of three distributions to the CARABAS II dataset images, namely, the Rayleigh, log-normal, and Rice distributions. It was particularly interesting, because of its broader selection of pdfs. It is important to note that [21] performs the statistical test on pixels along the stack dimension. Our focus is on the statistical representation of clutter along spatial dimensions, as discussed in the note at the end of Section III.

The Rayleigh distribution is suitable for modeling the amplitude of scattering regions in which the main contribution comes from additive noise and in which there is no prominent scatterer. The log-normal distribution has a nonnegative support and a heavy right tail. These characteristics make this distribution a suitable fit for histograms of radar intensity. The analysis in [21] shows that the log-normal distribution empirically fits the CARABAS II data well. The Rice distribution displays some of the same characteristics of the Rayleigh distribution, with the difference that it can model scattering regions where there is the contribution of prominent scatterers. The results in [21] verify that the Rice distribution can be a very good fit for wavelength-resolution SAR images.

As mentioned previously, the Gamma and Weibull will also be analyzed in this article. The Gamma distribution is used in the product model when the RCS is considered a constant [41]. The Gamma distribution can describe the characteristics of the RCS fluctuations of a heterogeneous terrain in high-resolution SAR images [42]. It is also used as the mean or modulation of sea radar echoes [43].

In [25], the use of the bivariate Gamma to model the magnitudes of wavelength-resolution SAR images is confirmed to be more realistic than the use of the bivariate Rayleigh distribution. The background clutter from high resolution or low grazing angle radar emission can be fit by the Weibull distribution by tuning its shape parameter [44]. This distribution can be used as an empirical model for the sea ice clutter [45], [46]. It has also been used to model the ground clutter [47], [48], including that of cultivated land [49].

Remark 1: The K distribution, which was mentioned in Section I as having been used in [39] for the study of the statistics of UWB VHF images, is not analyzed in this article because of its complex analytical expression. The resulting expressions for the entropy and the variance of the entropy are mathematically intractable. In this case, we conclude that the derivation of the entropy and the variance of the entropy are best conducted via approximations. In addition,

the ML parameter estimation for this distribution is much more complex than for the other distributions considered in this article. Hence, we reach a similar conclusion that the best alternative is to find function approximators for the parameter estimators. This is left for future work.

Table I lists the probability distributions for the CARABAS II dataset and their respective entropy and entropy variances. To easily follow Table I, some aspects of the distributions are presented as follows.

1) For the Gaussian distribution, we have

$$f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (10)$$

where μ and σ are the mean and standard deviation of X, respectively, and $\boldsymbol{\theta}^{\mathrm{T}} = \begin{pmatrix} \mu & \sigma \end{pmatrix}$.

2) For the log-normal distribution, we have

$$f(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \quad (11)$$

where μ and σ are the mean and standard deviation of $\ln(X)$, respectively, and $\boldsymbol{\theta}^{\mathrm{T}} = \begin{pmatrix} \mu & \sigma \end{pmatrix}$.

3) For the Rayleigh distribution, we have

$$f_X(x; \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (12)$$

where σ is the scale parameter or, alternatively, the standard deviation of the speckle noise, and $\boldsymbol{\theta} = \sigma$.

4) For the Gamma distribution, we have

$$f_X(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} \exp\left(-\frac{x}{\theta}\right) \quad (13)$$

where k and θ are the shape and scale parameters of the distribution, respectively, and $\boldsymbol{\theta}^{\mathrm{T}} = \begin{pmatrix} k & \theta \end{pmatrix}$. The function $\Gamma(z)$ is the standard Gamma function, $\Psi(z)$ is the Digamma function, and $\Psi_1(z)$ is the Trigamma function.

In the formulation of $\sigma_H^2(\boldsymbol{\theta})$, $\beta \triangleq 1 + (1-k)\Psi_1(k)$.

5) For the Weibull distribution, we have

$$f_X(x; \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right) \quad (14)$$

where λ and k are the shape and scale parameters, respectively, and $\boldsymbol{\theta}^{\mathrm{T}} = \begin{pmatrix} \lambda & k \end{pmatrix}$.

The constant γ is the Euler–Mascheroni constant.

The variables a_w, b_w, and c_w, which are used in the computation of the variance of the entropy $\sigma_H^2(\boldsymbol{\theta})$ for the Weibull pdf and which are featured in Table I, are presented in Appendix A for the sake of better readability.

6) The Rice distribution is given by

$$f_X(x; \nu, \sigma) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2+\nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right) \quad (15)$$

where ν and σ are the amplitude of the deterministic signal and the standard deviation of the speckle noise, respectively, and $\boldsymbol{\theta}^{\mathrm{T}} = \begin{pmatrix} \nu & \sigma \end{pmatrix}$.

TABLE I
ENTROPY AND THEIR VARIANCES FOR THE ANALYZED DISTRIBUTIONS

Distribution	Entropy H	Variance of the Entropy $\sigma_H^2(\theta)$
Gaussian	$\frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2)$	$1/2$
log-normal	$\frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2) + \mu$	$\sigma^2 + 1/2$
Gamma	$k + \ln(\theta) + \ln(\Gamma(k)) + (1-k)\Psi(k)$	$\left(k\beta^2 - 2\beta + \Psi_1(k)\right)/\left(k\Psi_1(k) - 1\right)$
Rayleigh	$1 + \ln\left(\frac{\sigma}{\sqrt{2}}\right) + \frac{\gamma}{2}$	$1/4$
Weibull	$\gamma\left(1 - \frac{1}{k}\right) + \ln\left(\frac{\lambda}{k}\right) + 1$	$[(b_w\gamma/k^2 - b_w/k - c_w/\lambda)(\gamma/k^2 - 1/k) + (1/\lambda)(-c_w\gamma/k^2 + c_w/k + a_w/\lambda)]/(a_w b_w - c_w^2)$
Rice	$1 + 2\ln(\sigma) + \frac{\nu^2}{\sigma^2} - \mathrm{E}_X[\ln(X)] - \mathrm{E}_X\left[\ln\left(I_0\left(\frac{\nu X}{\sigma^2}\right)\right)\right]$	$\left[b_R\left(\frac{\partial H}{\partial \nu}\right)^2 - 2c_R\frac{\partial H}{\partial \nu}\frac{\partial H}{\partial \sigma} + a_R\left(\frac{\partial H}{\partial \sigma}\right)^2\right] / \left(a_R b_R - c_R^2\right)$

Let $I_0(\cdot)$ be the modified Bessel function of the first kind with order 0 and $I_1(\cdot)$ be the modified Bessel function of the first kind with order 1.

Let the expected values featured in Table I be computed according to the Rice pdf of an RV X, as in (15).

We can observe in Table I the variables a_R, b_R, and c_R, as well as the partial derivatives $\frac{\partial H}{\partial \nu}$ and $\frac{\partial H}{\partial \sigma}$ of the entropy with respect to the parameters of the Rice pdf ν and σ, respectively. These variables and partial derivatives, which make up the expression of the variance of the entropy for the Rice distribution, are properly defined in Appendix B for the sake of readability.

V. CHANGE DETECTION ALGORITHM

In this section, we provide a description of the CDA we implemented. The pseudocode for the algorithm is presented in Algorithm 1.

A. Description

The CDA in Algorithm 1 can be divided into four distinct parts:
1) stack entropy statistic computation (lines 1–12);
2) median-based hypothesis testing (lines 13–15);
3) morphological operations (line 16);
4) computation of false alarm probability p_{FA} and detection probability p_D (line 17).

The stack entropy statistic is based on computing and comparing the entropy of all M images in each of the S stacks, for $M, S \in \mathbb{N}^*$. Let n_A be the number of elements in the azimuth direction and n_R be the number of elements in the range direction in a given 2-D grid.

To estimate distribution parameters via MLE, we perform a sliding window traversal of the image according to square-shaped regions of interest (ROI) of side length q centered around each pixel of the image, where $q \in \mathbb{N}^*$. Let $n_w \triangleq q^2$ be the number of pixels in each ROI, and let $\mathcal{R}$ be the set of all ROIs for a given image. For pixels on the edges, the ROI includes nondefined pixel positions. The edge pixels

can either be excluded from the analysis or be addressed using the same padding, i.e., by imputing the nondefined positions with the 0 value. With the same padding, we surround the original images with null values such that $|\mathcal{R}| = n_A n_R$. For the sake of simplicity, the remainder of the discussion on the CDA in Algorithm 1 assumes the same padding. However, the simulations and results presented in Section VI use valid padding, i.e., no padding. The only difference in valid padding is that

$$p_{\mathrm{proc}} \triangleq (n_A - q + 1)(n_R - q + 1)$$

pixels are processed, leaving

$$p_{\mathrm{exc}} \triangleq (q - 1)(n_A + n_R - q + 1) \tag{16}$$

edge pixels excluded from the analysis.

In particular, the element belonging to the ROI centered around the (i, j)th pixel, where $i \in \{1, \ldots, n_A\}$ and $j \in \{1, \ldots, n_R\}$, in image $\ell \in \{1, 2, \ldots, M\}$ and in stack $k \in \{1, 2, \ldots, S\}$ is represented by $x_{i,j,\ell,k}$. In some analyses, a computation will summarize information along the image axis, presenting a single statistics 2-D grid for the entire stack. In that case, individual grid elements have the corresponding notation $x_{i,j,k}$. Given that the choice of probability distribution affects the computation of the entropy (see Table I), we write in Algorithm 1 expressions based on the generic probability distribution of the image pixels $f_{X_{i,j,\ell,k}}(x_{i,j,\ell,k}; \theta_{i,j,\ell,k})$, where θ is the vector of distribution parameters estimated via MLE.

Entropy measurements are computed for each image according to the chosen pdf [(10), (11), (12), (13), (14), or (15)], its parameters θ estimated via MLE, and the Shannon differential entropy computed via (2). The mean of these entropy measurements is computed as in (7), and it is used in (8) to obtain one hypothesis test statistic $e_{i,j,k}$ in stack k for each (i, j)th pixel, where $k = 1, 2, \ldots, S$. Let $\mathbf{E}_k \triangleq (e_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$ be the matrix of all such statistics for stack k.

Entropy statistics from all S stacks are used to compute the elementwise stack-median entropy statistic matrix $\mathbf{E}_{\mathrm{med}}$, which is applied to each $\mathbf{E}_k$ via the masking operation $\mathcal{M}$,

Algorithm 1 CDA

$\{x_{i,j,\ell,k} : i \in \{1, 2, \ldots, n_A\}, j \in \{1, 2, \ldots, n_R\},$
$\ell \in \{1, 2, \ldots, M\}, k \in \{1, 2, \ldots, S\}\}$
Input:
Output: p_{FA}, p_D.

1: **for** $k = 1, 2, \ldots, S$ **do**
2:     **for** $\ell = 1, 2, \ldots, M$ **do**
3:         **for** $i = 1, 2, \ldots, n_A$ **do**
4:             **for** $j = 1, 2, \ldots, n_R$ **do**
5:                 Estimate $\boldsymbol{\theta}_{i,j,\ell,k}$ via MLE.
6:                 $h_{i,j,\ell,k} = -\int_{-\infty}^{\infty} f_{X_{i,j,\ell,k}}(x_{i,j,\ell,k}; \boldsymbol{\theta}_{i,j,\ell,k})$
                              $\ln(f_{X_{i,j,\ell,k}}(x_{i,j,\ell,k}; \boldsymbol{\theta}_{i,j,\ell,k})) \, dx_{i,j,\ell,k}.$
7:                 $\mu_{i,j,\ell,k} = \int_{-\infty}^{\infty} x_{i,j,\ell,k}$
                              $f_{X_{i,j,\ell,k}}(x_{i,j,\ell,k}; \boldsymbol{\theta}_{i,j,\ell,k}) \, dx_{i,j,\ell,k}.$
8:                 $\sigma_{i,j,\ell,k}^2 = \int_{-\infty}^{\infty} (x_{i,j,\ell,k} - \mu_{i,j,\ell,k})^2$
                              $f_{X_{i,j,\ell,k}}(x_{i,j,\ell,k}; \boldsymbol{\theta}_{i,j,\ell,k}) \, dx_{i,j,\ell,k}.$
9:         Create the entropy matrix
        $\mathbf{H}_{\ell,k} = (h_{i,j,\ell,k})_{1 \le i \le n_A, 1 \le j \le n_R}.$
10:         Create the matrix with the reciprocal of the variances:
        $\mathbf{R}_{\ell,k} = (1/\sigma_{i,j,\ell,k}^2)_{1 \le i \le n_A, 1 \le j \le n_R}.$
11:         Create the stack mean-entropy $\overline{\mathbf{H}}_k = \frac{1}{M} \sum_{\ell=1}^{M} \mathbf{H}_{\ell,k}.$
12:         $\mathbf{E}_k = \sum_{\ell=1}^{M} N_{\ell,k}[\mathbf{R}_{\ell,k} \odot (\mathbf{H}_{\ell,k} - \overline{\mathbf{H}}_k)^{\odot 2}],$
        where $N_{\ell,k}$ is the size of the windows in image ℓ and stack k (in our case, $N_{\ell,k}$ is constant through all the images of all stacks and equal to n_w), $\mathbf{W}^{\odot 2}$ is the elementwise squaring of matrix $\mathbf{W}$, and $\odot$ is the Hadamard matrix multiplication operator.
13: Compute the elementwise median $\mathbf{E}_{\text{med}}$ of the vector of matrices $(\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_S)$.
14: Mask the values $\mathbf{E}_k$ via the Hadamard product with $\mathbf{E}_{\text{med}}$: $\mathbf{M}_k \triangleq \mathcal{M}(\mathbf{E}_k, \mathbf{E}_{\text{med}}) = \mathbf{E}_k \odot \mathbf{E}_{\text{med}}$, for $k = 1, 2, \ldots, S$.
15: Let $\mu_{i,j,k}$ be the (i, j)-element of $\mathbf{M}_k$ such that $\mathbf{M}_k = (\mu_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$. Compute

$\mathbf{T}_k \triangleq (t_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$
$= (\mathbf{1}_{\mathfrak{T}_k}(\mu_{i,j,k}))_{1 \le i \le n_A, 1 \le j \le n_R} \in \{0, 1\}^{n_A \times n_R},$

    where $\mathfrak{T}_k = \{x \in \mathbb{R} : x > T_k\}$ for an empirically determined threshold $T_k \in \mathbb{R}$ and where $\mathbf{1}_{\mathfrak{T}_k}$ is the indicator function according to $\mathfrak{T}_k$.
16: Perform morphological operations (one erosion and two dilations) over all $n_A n_R$ elements $t_{i,j,k}$ of $\mathbf{T}_k$: $\mathbf{A}_k = (\mathcal{A}(t_{i,j,k}))_{1 \le i \le n_A, 1 \le j \le n_R}$, where $\mathcal{A}$ is the elementwise morphological map.
17: Compute p_{FA} and p_D for every element of $\mathbf{A}_k$ according to Remark 2.

---

which concretely performs the mapping

$$\mathcal{M}(\mathbf{E}_k, \mathbf{E}_{\text{med}}) : \mathbf{E}_k \mapsto \mathbf{E_k} \odot \mathbf{E}_{\text{med}}$$

that is, the Hadamard product between $\mathbf{E}_k$ and $\mathbf{E}_{\text{med}}$. This product creates matrix $\mathbf{M}_k \triangleq \mathbf{E}_k \odot \mathbf{E}_{\text{med}} = (\mu_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$.

The elements $\mu_{i,j,k}$ undergo a thresholding operation $\tau : \mathbb{R} \to \{0, 1\}$ defined as

$$t_{i,j,k} \triangleq \tau(\mu_{i,j,k}) = \mathbf{1}_{\mathfrak{T}_k}(\mu_{i,j,k})$$

where $\mathfrak{T}_k = \{x \in \mathbb{R} : x > T_k\}$ and $T_k \in \mathbb{R}$ is an empirically determined threshold. The function $\mathbf{1}_{\mathfrak{T}_k}(\cdot)$ is the indicator function according to $\mathfrak{T}_k$. Let $\mathbf{T}_k$ be the matrix that contains all $n_A n_R$ $t_{i,j,k}$ values for stack k such that $\mathbf{T}_k \triangleq (t_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$. In what follows, we will abuse notation and use the operation τ also for matrices, namely $\mathbf{T}_k = \tau(\mathbf{M}_k)$.

After masking and thresholding, the processed statistics undergo a series of elementwise morphological operations, including an $f \times f$ erosion and two $g \times g$ dilations, where f and g are the integer side lengths of the square kernels utilized in erosion and dilation, respectively. Let the concatenation of the morphological operations be represented by $\mathcal{A}$, where $\mathcal{A}$ maps elements $t_{i,j,k}$ to $\alpha_{i,j,k}$: $\alpha_{i,j,k} = \mathcal{A}(t_{i,j,k})$. Let $\mathbf{A}_k$ be the matrix formed by all such $\alpha_{i,j,k}$ such that $\mathbf{A}_k = (\alpha_{i,j,k})_{1 \le i \le n_A, 1 \le j \le n_R}$. We will again abuse notation and expand the definition of $\mathcal{A}$ to matrices such that $\mathbf{A}_k = \mathcal{A}(\mathbf{T}_k)$.

Let the RV corresponding to the individual elements in $\mathbf{E}$ matrices be denoted by $E_{i,j,k}$. This RV is distributed according to $\chi^2_{n_w-1}$. The composition of the morphological operations $\mathcal{A}$ with thresholding τ and the nonlinear mask $\mathcal{M}$ maps $E_{i,j,k}$ to the RV

$$A_{i,j,\ell,k} = \mathcal{A} \circ \tau \circ \mathcal{M}(E_{i,j,k}, E_{i,j,\text{med}})$$

whose distribution is $f_{A_{i,j,k}}(\alpha_{i,j,k})$. The value $\alpha_{i,j,k}$ divides the support of $E_{i,j,k}$ into the partitions $\mathcal{P}_0$ and $\mathcal{P}_1$. When $\alpha_{i,j,k} = 0$, we have the scenario of the null hypothesis $\mathcal{H}_0$ in (4), which corresponds to $\mathcal{P}_0$, whereas when $\alpha_{i,j,k} = 1$, we have the scenario of the alternative hypothesis $\mathcal{H}_1$ in (5), which corresponds to $\mathcal{P}_1$:

$$\mathcal{P}_0 = \{e_{i,j,k} : \alpha_{i,j,k} = \mathcal{A} \circ \tau \circ \mathcal{M}(e_{i,j,k}, e_{i,j,\text{med}}) = 0\}$$
$$\mathcal{P}_1 = \{e_{i,j,k} : \alpha_{i,j,k} = \mathcal{A} \circ \tau \circ \mathcal{M}(e_{i,j,k}, e_{i,j,\text{med}}) = 1\}.$$

We theoretically define the probabilities of detection p_D and false alarm p_{FA} for each element $\alpha_{i,j,k}$ as

$$p_D = p_D(\alpha_{i,j,k}) \triangleq \int_{\mathcal{P}_0} f_{A_{i,j,k}}(\alpha_{i,j,k}) \, d\alpha_{i,j,k} \tag{17}$$

$$p_{FA} = p_{FA}(\alpha_{i,j,k}) \triangleq \int_{\mathcal{P}_1} f_{A_{i,j,k}}(\alpha_{i,j,k}) \, d\alpha_{i,j,k}. \tag{18}$$

Remark 2: In practice, p_{FA} and p_D are computed differently from (17) and (18). We binarize the elements of $\mathcal{M}(\mathbf{E}_k)$ by thresholding them with an empirically chosen value T_k. The output of binarization has no analytical probability distribution formula. It consists on assigning to all elements of $\mathcal{M}(\mathbf{E}_k)$ logical true or false values. We overlay the output of this binary mapping with a map of ground truths and further consider morphological criteria, such as pixel size and layout of pixel clusters, to distinguish between detected targets and false alarms. The count of events matching the ground truths divided by the total number of events results in the p_D, whereas the count of mismatches divided by the total number of events results in the p_{FA}.

Variations to Algorithm 1 are due to the absence of median computation or morphological operations. The version presented in Algorithm 1 is the most complete. To derive alternative algorithms, it is possible to leave out lines 9–11

for an algorithm without median computation or line 13 for an algorithm without morphological operations.

In Fig. 2(a), we show the flowchart for Algorithm 1 without the masking operation. This flowchart includes parts 1), 3), and 4), which were discussed at the beginning of this subsection. Namely, the following operations are depicted: the computation of the entropy-based statistic, morphological operations, and the computation of false alarm and detection probabilities.

The operations in Fig. 2(a) are described pixelwise. The (i, j)th pixel of the M images composing stack k, $k \in \{1, 2, \ldots, S\}$, is used in the CDA, where $i \in \{1, 2, \ldots, n_A\}$ and $j \in \{1, 2, \ldots, n_R\}$. The entropy $H_\ell(\hat{\boldsymbol{\theta}}_\ell)$, $\ell \in \{1, 2, \ldots, M\}$, of pixel (i, j) is computed for each image according to (2). Next, entropies $H_\ell(\hat{\boldsymbol{\theta}}_\ell)$ are used in the computation of the entropy-based statistic, as given by (9). The entropy-based statistic e is then thresholded by the threshold operator τ. The output t is subjected to the morphological operations of erosion and dilation represented by mapping $\mathcal{A}$, yielding signal α. This signal then becomes the input to the probability functionals $\mathbb{P}_D$ and $\mathbb{P}_{FA}$ in (17) and (18), respectively.

In Fig. 2(b), the operations following the computation of the entropy-based statistics are shown with the inclusion of the masking mapping of part 2) at the beginning of this subsection. We show in the flowchart that the input to masking $\mathcal{M}$ consists of the entropy-based statistics $(e_1, e_2, \ldots, e_S)$, which are, respectively, computed with the (i, j)th pixels of the images making up each one of the S different stacks used in Algorithm 1. The entropy-based statistics for each of the S stacks $(e_1, e_2, \ldots, e_S)$ are used to compute the median statistic, with which median masking $\mathcal{M}$ is performed. Similar to the output e in Fig. 2(a), the outcome μ of median masking is then capped and binarized by a threshold operator τ. The output t then undergoes the morphological operations of erosion and dilation represented by mapping $\mathcal{A}$ to produce signal α. Signal α is then used in the computation of the probability of detection (PD) p_D in (17) and of the probability of false alarm p_{FA} in (18).

For a given stack with M images, change detection occurs as if the scheme compares one surveillance image with $M - 1$ reference images for any given change detection operation and if M such operations occur per stack. It is important to note, though, that in practice, all targets from all missions are detected through a single hypothesis test. It is as if one single change image were created from the stack, detecting the changes among all stack images with one set of statistics.

B. Complexity

For the computation of the entropy and its variance, it can be seen from Table I that once parameters have been found through MLE (step 5 of Algorithm 1), the complexity is at most linear in the number of pixels $n_A n_R$ of the image, i.e., $\mathcal{O}(n_A \times n_R)$. ML estimation and entropy-related computations scale linearly with the number of images M and also with the number of stacks N. Hence, the overall complexity, without taking the MLE into account, is $\mathcal{O}(N \times M \times n_A \times n_R)$.

Since each probability distribution has its own defining pdf, the numerical complexity of the MLE parameter estimation can vary for each distribution. Henceforward, we discuss the differences between the numerical complexity of the MLE for the distributions mentioned in Table I.

1) Gaussian: To find the parameter estimators $\hat{\mu}$ and $\hat{\sigma}$ based on MLE for the Gaussian distribution in (10), we use log-likelihood function

$$\ell(\mu, \sigma) = \sum_{i=1}^{n} \left[-\frac{\ln(2\pi)}{2} - \ln \sigma - \frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2 \right].$$

To estimate the MLE parameters, we calculate the partial derivatives of the log-likelihood function with respect to μ and σ and set them to 0. This yields the estimators

$$\hat{\mu} = \frac{1}{n_w} \sum_{i=1}^{n_w} x_i \tag{19}$$

$$\hat{\sigma}^2 = \frac{1}{n_w} \sum_{i=1}^{n_w} x_i^2 - \frac{1}{n_w^2} \left(\sum_{i=1}^{n_w} x_i\right)^2. \tag{20}$$

The calculations in (19) and (20) require sums over $\mathcal{O}(1)$ operations performed on the n_w data points of the sliding window and an $\mathcal{O}(1)$ operation over the sum of these data points. Hence, the time complexity is $\mathcal{O}(n_w)$.

The overall complexity for the computation of the statistic in (6) under the Gaussian distribution is $\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$.

2) Log-Normal: For the log-normal distribution, given by the pdf in (11), we formulate the log-likelihood function as follows:

$$\ell(\mu, \sigma) = \sum_{i=1}^{n} \left[-\ln\left(x_i \sqrt{2\pi\sigma^2}\right) - \frac{(\ln(x_i) - \mu)^2}{2\sigma^2} \right].$$

To determine the MLE parameters, we again compute the partial derivatives of the log-likelihood function with respect to μ and σ and equate them to 0. Consequently, the estimators

$$\hat{\mu} = \frac{1}{n_w} \sum_{i=1}^{n_w} \ln x_i \tag{21}$$

$$\hat{\sigma}^2 = \frac{1}{n_w} \sum_{i=1}^{n_w} \left(\ln x_i - \hat{\mu}\right)^2 \tag{22}$$

are obtained.

The evaluations in (21) and (22) involve sums over $\mathcal{O}(1)$ operations conducted on the n_w data points of the sliding window, as well as $\mathcal{O}(1)$ operations over the aggregate of the logarithms of these data points. As a result, the time complexity is $\mathcal{O}(n_w)$.

As for the Gaussian distribution, the total complexity for calculating the statistic under the log-normal distribution is $\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$.

3) Rayleigh: For the Rayleigh distribution, given by the pdf in (12), we write the log-likelihood function as

$$\ell(\sigma) = \sum_{i=1}^{n} \left[\ln x_i - \ln \sigma^2 - \frac{x_i^2}{2\sigma^2} \right].$$
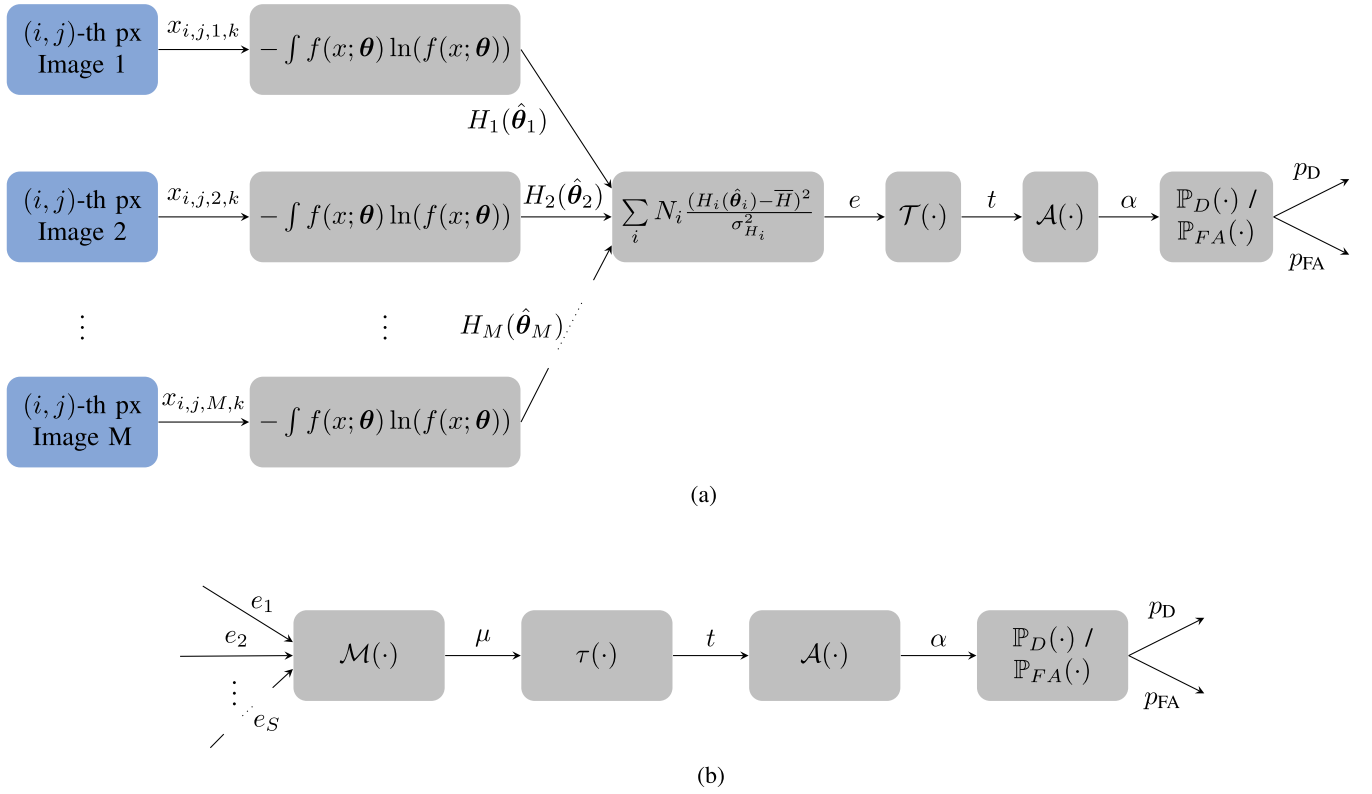
(a)

(b)

Fig. 2. Flowcharts for the CDA algorithm with and without masking. (a) Flowchart for the change detection without median masking. (b) Flowchart with the inclusion of median masking after the entropy-based statistics.

To proceed with the MLE, the derivative of the log-likelihood function with respect to σ is, as before, calculated and set to 0. This yields the estimator

$$\hat{\sigma}^2 = \frac{1}{2n_w} \sum_{i=1}^{n_w} x_i^2. \tag{23}$$

The calculations in (23) require sums over $\mathcal{O}(1)$ operations performed on the n_w data points of the sliding window. Hence, the time complexity is $\mathcal{O}(n_w)$.

The overall complexity for the computation of the statistic under the Rayleigh distribution is $\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$.

4) *Gamma:* For the Gamma distribution, defined by the pdf in (13), we construct the log-likelihood function as follows:

$$\ell(k, \theta) = \sum_{i=1}^{n} \left[-(k-1)\ln x_i - \ln \Gamma(k) - \frac{x_i}{\theta} - k \ln \theta \right].$$

Again, we compute the partial derivatives of the log-likelihood function concerning k and θ and set them to 0. To solve the two resulting equations, it is generally necessary to employ numerical methods, as closed-form expressions cannot be obtained. For this reason, the MLE process can have an arbitrary complexity based on the chosen method for parameter estimation. Formally, after setting a solving method, the complexity of the process can be described as $\mathcal{O}[f_\gamma(n_w, \alpha)]$, where α is a vector of the parameters of the chosen method for parameter estimation, such as the Newton–Raphson method, and $f_\gamma(n_w, \alpha)$ is the function that maps n_w and α to the numerical complexity of the MLE process for the Gamma distribution. Thus, the overall complexity

for computing the statistic under the Gamma distribution is $\mathcal{O}[N \times M \times n_A \times n_R \times f_\gamma(n_w, \alpha)]$.

5) *Weibull:* Following the pdf in (14), the log-likelihood function is as follows:

$$\ell(\lambda, k) = \sum_{i=1}^{n} \left[\ln k - \ln \lambda + (k-1)\ln\left(\frac{x_i}{\lambda}\right) - \left(\frac{x_i}{\lambda}\right)^k \right].$$

The parameters λ and k can be estimated again by setting their partial derivatives to zero. As for the Gamma case, solving the resulting equations typically requires numerical methods. Thus, again, the MLE process can exhibit an arbitrary complexity depending on the chosen method for parameter estimation, and the complexity can be expressed as $\mathcal{O}[f_\omega(n_w, \alpha)]$, where α is a vector of the parameters of the chosen method for parameter estimation, and $f_\omega(n_w, \alpha)$ is the function that maps n_w and α to the numerical complexity of the MLE process for the Weibull distribution. Consequently, the overall complexity for computing the statistic under the Weibull distribution is $\mathcal{O}[N \times M \times n_A \times n_R \times f_\omega(n_w, \alpha)]$.

6) *Rice:* For the Rice distribution, with the pdf given by (15), the log-likelihood function can be expressed as

$$\ell(v, \sigma) = \sum_{i=1}^{n} \left[\ln\left(\frac{x_i}{\sigma^2}\right) - \frac{x_i^2 + v^2}{2\sigma^2} + \ln I_0\left(\frac{x_i v}{\sigma^2}\right) \right].$$

Similar to the previous cases, finding the solution for the resulting equations when the partial derivatives of the log-likelihood function are set to zero typically requires numerical methods. As for the two previous distributions, the MLE process exhibits an arbitrary complexity depending on

TABLE II

TIME COMPLEXITY FOR THE ANALYZED DISTRIBUTIONS AND AVERAGE ELAPSED TIME FOR THE EXECUTION OF THE PROPOSED ALGORITHM WITHOUT MEDIAN MASKING

Distribution	Time Complexity	Elapsed Time (s)
Gaussian	$\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$	103
Log-Normal	$\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$	148
Rayleigh	$\mathcal{O}(N \times M \times n_A \times n_R \times n_w)$	56
Gamma	$\mathcal{O}[N \times M \times n_A \times n_R \times f_\gamma(n_w, \alpha)]$	5381
Weibull	$\mathcal{O}[N \times M \times n_A \times n_R \times f_\omega(n_w, \alpha)]$	N/A
Rice	$\mathcal{O}[N \times M \times n_A \times n_R \times f_\rho(n_w, \beta)]$	N/A

the chosen method for parameter estimation, and, similarly, the complexity can be expressed as $\mathcal{O}[f_\rho(n_w, \beta)]$, where β is a vector of the parameters of the chosen method for parameter estimation, and $f_\rho(n_w, \beta)$ is the function that maps n_w and β to the numerical complexity of the MLE process for the Rice distribution. Consequently, the overall complexity for computing the statistic under the Rice distribution is $\mathcal{O}[N \times M \times n_A \times n_R \times f_\rho(n_w, \beta)]$.

A summary of the time complexity results can be found in Table II, where we present in the second column the time complexity expressions for all distributions analyzed. In the third column, we present the average time complexity in seconds when running a single-stack detection, i.e., running Algorithm 1 with $S = 1$, without median masking. Results are available for the Gaussian, Log-Normal, Rayleigh, and Gamma probability distributions, which are those that permit an analysis within feasible time ranges. These numbers were obtained by running the process multiple times for each pdf (in order to average out statistical fluctuations) on a machine equipped with 16 GB RAM DDR4 memory, powered by an Apple M2 ARM processor containing eight physical cores, and running at a standard clock frequency of 3.5 GHz.

In Table II, we can observe that the time complexity for the Gamma, Weibull, and Rice distributions contain the factors $f_\gamma(n_w, \alpha)$, $f_\omega(n_w, \alpha)$, and $f_\rho(n_w, \beta)$, respectively. By the fact that these factors are related to iterative methods for the ML estimation of distribution parameters, which are usually based on root-finding procedures such as the Newton–Raphson method, they are almost always greater than n_w. In the case of the Weibull and Rice distributions, their respective factors tend to be much greater than n_w, to the point that our computation resources were not sufficient for the ML estimation of the parameters of these distributions in a feasible amount of time.

We also see in Table II that the Gaussian, Log-Normal, and Rayleigh distributions have running times in the same order of magnitude. This verifies the claim made in the second

column of Table II that their time complexities are equal up to multiplicative constants.

In spite of the Gamma distribution's need for iterative parameter estimation methods, we were able to produce running time results for it. These results evidence an elapsed processing time orders of magnitude longer than that observed for the Gaussian, Log-Normal, and Rayleigh distributions. This disparity is again due to the ML estimation process, which is iterative, thus usually involving many more mathematical operations than those required for estimation methods for which we have closed analytical expressions, such as is the case of the Gaussian, Log-Normal, and Rayleigh pdfs.

VI. SIMULATIONS AND RESULTS

We present in this section simulation results considering our CDA applied to the CARABAS II dataset. The CDA used is the full version of Algorithm 1.

For all images, $n_A = 3000$ pixels and $n_R = 2000$ pixels. The length of the side of the square sliding window is $q = 11$. This choice of length is empirical and based on the best compromise found between the goodness-of-fit of distributions regarding data histograms obtained for each q^2-sized window centered around each valid pixel of the image and computational cost. Our processing uses valid padding, as described in Section V. From (16), we exclude $p_{exc} = 49,900$ edge pixels from the analysis, i.e., approximately 0.832% of the pixels are excluded.

The two metrics used to measure performance are the PD and the FAR. The PD, defined in (17), is computed as described in Remark 2. The FAR is related to the p_{FA} defined in (18), with the caveat that it substitutes the denominator in Remark 2 by the area of the whole analysis region (in this case, the entire 6 million square pixels of the image, which correspond to 6 km², as each pixel has a size of 1×1 m).

Our process employs morphological operations. Following the naming conventions from Section V-A, we choose an erosion kernel of size $f = 3$ and dilation kernels of size $g = 3$. The choices of f and g result from an experimental minimax procedure attempting to minimize the FAR for the maximum possible PD.

In possession of the PD and the FAR, we compute receiver operating characteristic (ROC) curves [50], and we compare our results with results from the state of the art [18], [19], [20], [21]. The area under curve (AUC) values for such curves are used as a performance indicator.

The change detection analysis in Algorithm 1 uses either pass-based (PB) or flight-geometry-based (FGB) stacks.

1) PB: $S = 6$ stacks composed of $M = 4$ images selected from the same pass, i.e., with the same flight angle.
2) FGB: $N = 3$ stacks composed of $M = 8$ images selected from the same flight geometry.

The distributions of Section IV are used to model the statistics of the clutter-plus-noise. In Fig. 3, we show the ROC curves for our CDA algorithm for PB stacks without median masking. In Fig. 4, we show the ROC curves for our CDA algorithm without median masking but for FGB stacks. It can be seen that performance for PB and FGB stacks can be
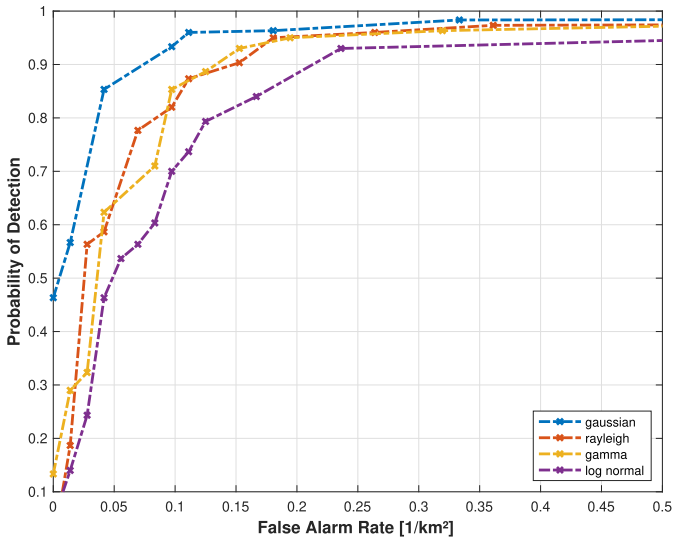
(a)
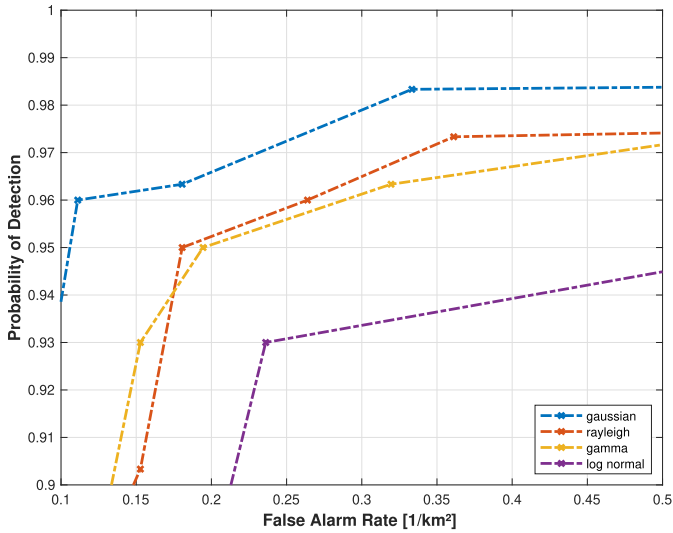
(b)

Fig. 3. Receiver operating characteristic curves obtained from simulations of the CDA in Algorithm 1 using PB stacks without median masking for the Gaussian, Rayleigh, Gamma, and log-normal clutter-plus-noise distributions. (a) Expanded view with all simulated points of operation in terms of PD and FAR. (b) Zoomed-in view at the elbow of the curves to highlight the transition into the highest possible PD for each distribution.

TABLE III
AREA UNDER THE SIMULATED ROC CURVES (AUC) FOR EACH SELEC-TION OF PROBABILITY DISTRIBUTION FITTING USING FGB AND PB STACKS, WITHOUT MEDIAN MASKING. THE AREA IS CALCU-LATED FOR THE FAR INTERVAL [0, 0.5]

Probability Distribution	AUC Pass-based	AUC Geometry-based
Gaussian	0.469	0.480
Rayleigh	0.439	0.464
Gamma	0.434	0.461
log-normal	0.402	0.450
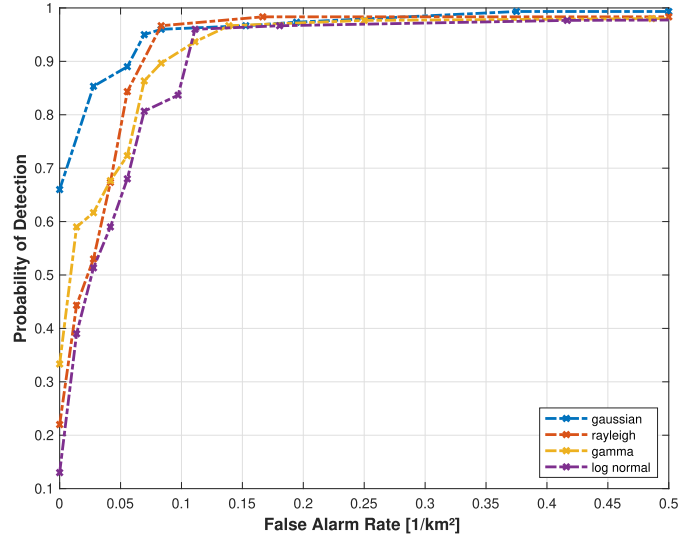
(a)

(b)

Fig. 4. Receiver operating characteristic curves obtained from simulations of the CDA in Algorithm 1 using FGB stacks without median masking for the Gaussian, Rayleigh, Gamma, and log-normal clutter-plus-noise distributions. (a) Expanded view with all simulated points of operation in terms of PD and FAR. (b) Zoomed-in view at the elbow of the curves to highlight the transition into the highest possible PD for each distribution.

ordered from highest to lowest performing clutter-plus-noise distribution: Gaussian, Rayleigh, Gamma, and log-normal. Table III displays the AUC values for the ROC curves in Figs. 3 and 4, i.e., for all PB and FGB stacks, respectively, without median masking.

In Fig. 5, ROC curves are plotted for the distributions in Section IV using median masking with both PB and FGB stacks. Fig. 5(a) focuses on PB stacks, while Fig. 5(b) provides the corresponding results for FGB stacks. We notice from these figures that the use of median masking has a noticeable improvement in terms of the increase in PD and decrease in FAR on the ROC curves in comparison with the curves from Figs. 3 and 4. Another feature that stands out is that the results of PB stacks are almost always better than those of FGB stacks, as it can be verified by means of Table IV,
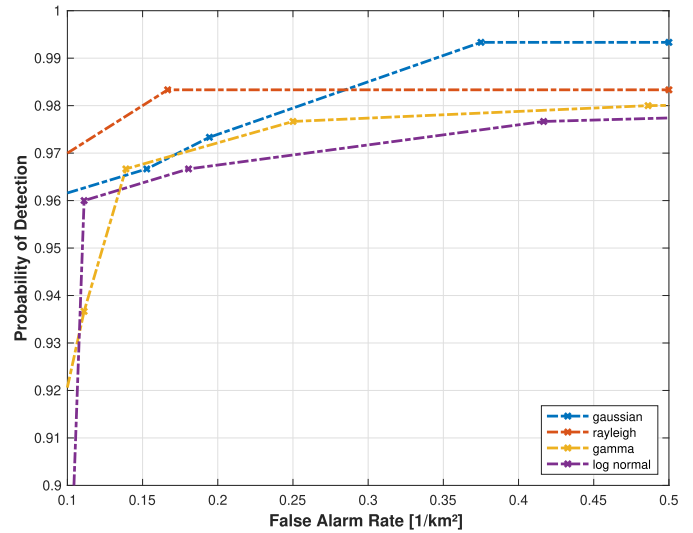
which shows AUC values for PB and FGB stacks under the effect of median masking.
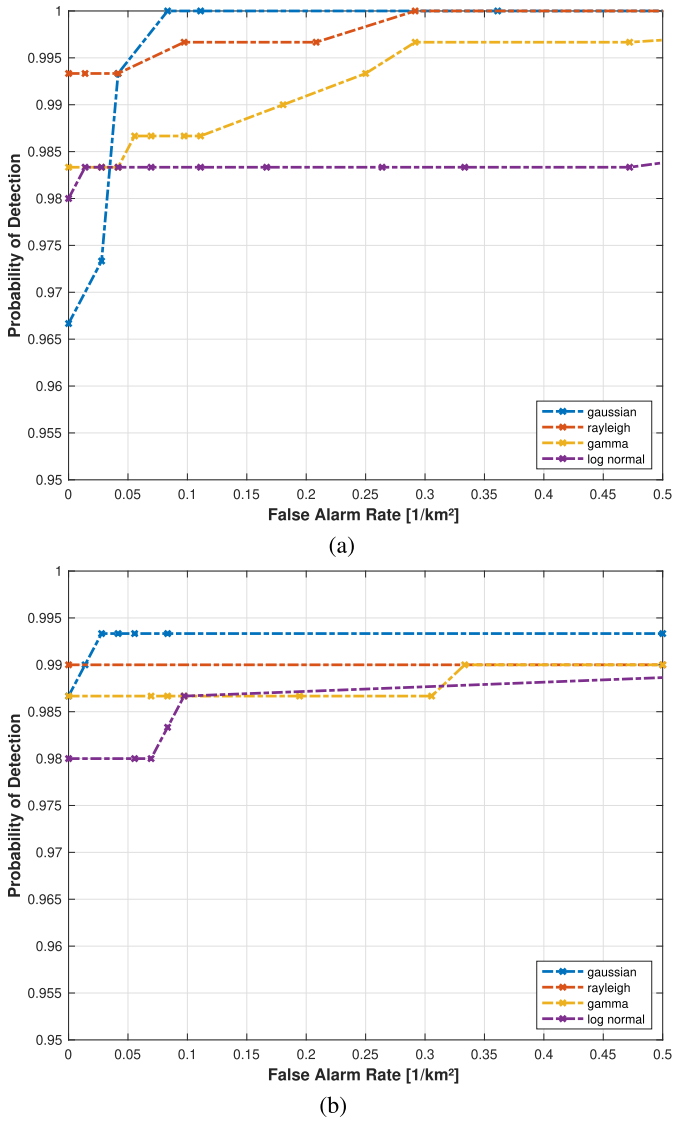
(a)

(b)

Fig. 5. Receiver operating characteristic curves obtained from simulations of the CDA in Algorithm 1 with median masking for the Gaussian, Rayleigh, Gamma, and log-normal clutter-plus-noise distributions. (a) Curves created using PB stacks. (b) Curves created using FGB stacks.

TABLE IV
AREA UNDER THE SIMULATED ROC CURVES (AUC) FOR EACH SELECTION OF PROBABILITY DISTRIBUTION FITTING USING FGB AND PB STACKS, WITH MEDIAN MASKING. THE AREA IS CALCULATED FOR THE FAR INTERVAL $[0, 0.5]$

Probability Distribution	AUC Pass-based	AUC Geometry-based
Gaussian	0.499	0.497
Rayleigh	0.499	0.495
Gamma	0.496	0.494
log-normal	0.492	0.493

A. Discussion

In the analyses with and without median masking, the general trend is for the Gaussian to show the best results and then the Rayleigh, Gamma, and log-normal, respectively. The first conjecture is that this is also the order of best spatial distribution fits. At a deeper level, the difference in performance among distributions could be due to the variables on which the entropy and its variance depend. As can be seen in Table I, the Gaussian entropy depends solely on the variance σ^2, the Rayleigh on its scale parameter, the Gamma on the shape parameter k and on the scale parameter θ, and the log-normal on the mean μ and the standard deviation of the RV's natural logarithm σ. Based on that, the second conjecture is that the entropy function of the image's sliding windows has the most mutual information with the variance of the Gaussian distribution or the scale parameter of the Rayleigh, as opposed to the parameters of the Gamma and the log-normal. Since the Gamma and the log-normal have entropy functions depending on two parameters, another possibility is that these functions have different dependences on each parameter and are positively correlated to one and negatively to the other for certain parameter values in the parameters' value space. Most importantly, and now looking at the variance of the entropy in Table I, the variances for the Gaussian and Rayleigh are constant, while it depends on distribution parameters for the log-normal and Gamma. Hence, alterations in parameters from window to window could make the entropy of the Gamma and log-normal have more variability around the expected value than their Gaussian and Rayleigh counterparts. The interplay between distributions and entropy, particularly their parameters, is a topic to be studied in a future article.

The median masking makes some implicit assumptions. It assumes that the statistics of images of the same scene for the same targets but from different flight geometries differ in terms of the false alarms presented. Given that stacks in this article are built with regard to pass and flight geometry, the previous assumption is then extended to the stacks, in particular to stack entropy-based statistics. The assumption seems reasonable in light of the different angles used. The median serves as an anomaly detector, excluding artifacts that appear in certain angles and not in others and, by extension, in certain stacks but not in others. In fact, any other robust statistic could have been used in place of the median. The choice for the median is grounded on the work of [18] and [27]. It is also based on the work of [21] on the statistics of VHF-UWB images in the stack dimension. Since the best fitting distributions in this dimension seem to be those that have long tails, again, the choice for robust statistics as a masking device seems justified. The choice of mask made in this article is highly dependent on the changes in geometry among stacks, but it seems relatively general from the point of view of change detection in other scenarios, as the radar imagery of swaths of terrain is very likely to be acquired at different geometries among the different acquisition times. All the hypotheses raised in this paragraph will be researched in a future article.

The implementation of median masking significantly lowers the FAR across all distribution types, as demonstrated in Figs. 3 and 4. In Fig. 6, the false alarms for the usage of entropy-based statistic matrices (already subjected to an empirical threshold) observed in Fig. 6(a) for the Rayleigh distribution and the FGB stack are suppressed by median masking in (d). By employing the stack median, the impact of elongated structures is mitigated, consequently reducing
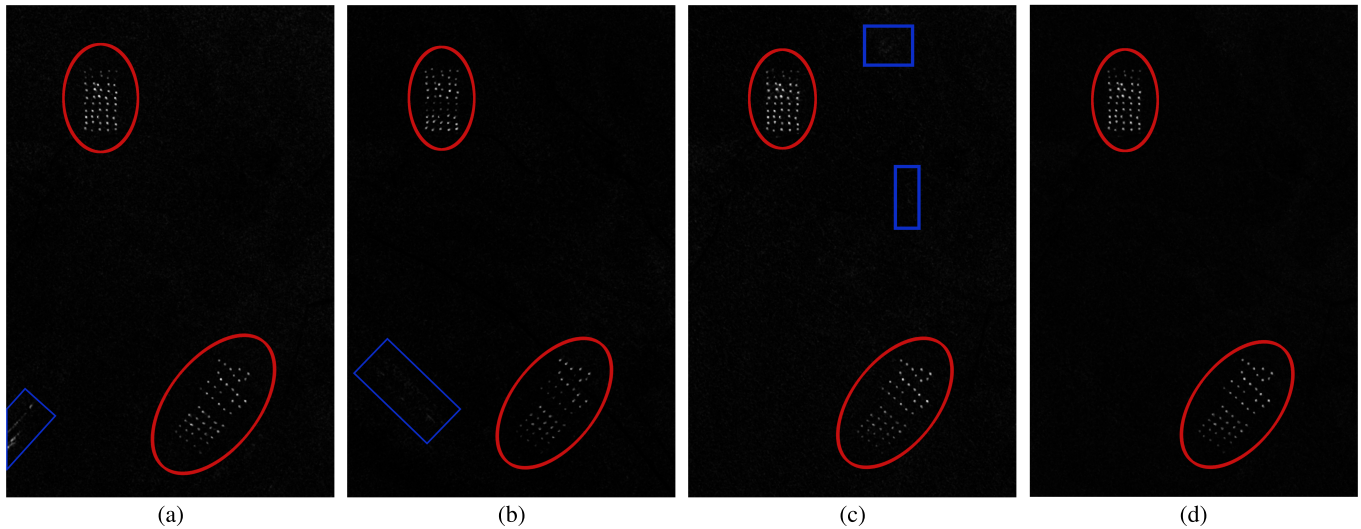
Fig. 6. (a)–(c) Pictorial representation of entropy-based statistic matrices obtained by means of Algorithm 1 for the Rayleigh distribution using FGB stacks. (d) End result after the application of median masking in (d). (a)–(d) Red ellipses refer to ground-truth targets, while blue rectangles refer to artifacts capable of producing false alarms. (a) Stack 1: Passes 1 and 3. (b) Stack 2: Passes 2 and 4. (c) Stack 3: Passes 5 and 6. (d) Median masking.

the likelihood of false alarms. In Fig. 6, we can observe that Fig. 6(a)–(c) contain artifacts that may be interpreted as false alarms. The most prominent of all belongs to Stack 1 in Fig. 6(a), in the bottom left corner of the picture. The artifacts of Stack 2 in Fig. 6(b) are less detectable in terms of intensity, but they cover a wider area. The artifacts of Stack 3 are the least detectable and occupy a relatively small area. Most importantly, though, the different flight geometries of Stacks 1, 2, and 3 ensure that the artifacts appear in different positions of the frame. Different flight geometries produce distinct reflectance characteristics, which in turn usually lead to alterations in the configuration of the entropy-based statistic matrices. The matrices change but conserve similarities mainly in the regions of highest intensity, which, in this case, are the targets that can, to a larger or lesser degree, be detected from all flight angles.[2] The benefits of the lack of overlap among artifacts from different flight geometries are that artifacts of any given stack are dampened by the median entropy-based statistic computed over all stacks, thus reducing the risk of false alarms. This phenomenon can be observed in Fig. 6(d), which shows clear signs of smoothing outside of the target regions.

In the absence of median masking, a minor performance improvement can be observed with flight geometry-based stacking compared to PB stacking. Generally, PB stacks exhibit similar FARs to flight-geometry-based stacks. This can be attributed to the fact that the four images comprising the image stacks in PB selection provide sufficient information to suppress most potential false alarms. As a result, adding more images to a stack with analogous characteristics from the same scene may yield diminishing performance returns.

[2]Sensors will not always pick up the targets, as, depending on their nature, they may be difficult to detect. However, we assume that sensor design and flight geometry selection are optimally tailored for the desired targets. Hence, we suppose that most, if not all, flight geometries selected by a measurement campaign will be able to detect a large portion of the targets.
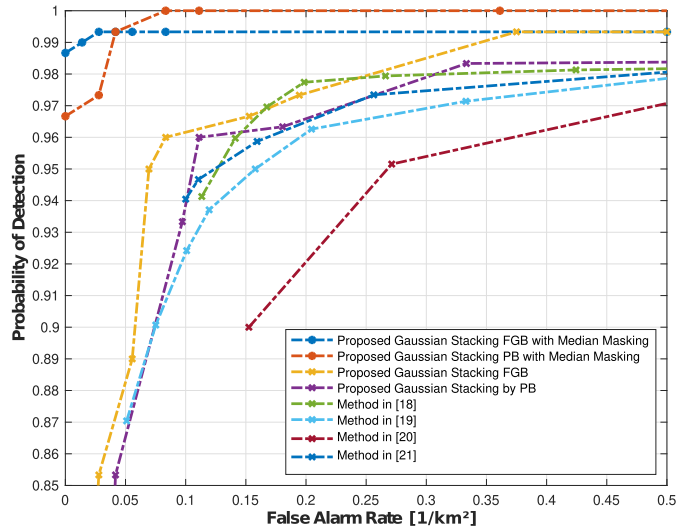

Fig. 7. Comparative performance of the state-of-the-art and proposed algorithms with and without median masking.

Interestingly, when median masking is applied, PB stacking displays a marginally higher AUC than FGB stacking. This could be explained by both cases approaching the maximum AUC of 0.5, rendering the 0.02 difference in their AUCs potentially statistically insignificant. Additionally, Fig. 7 contrasts the ROC of FGB and PB stacking with Gaussian fitting, both with and without median GSP masking, revealing an overall enhancement over the results presented in [18], [19], and [21]. At a specific operating point, the Gaussian distribution without ground scene median masking achieves FAR = $0.08/\text{km}^2$ for $P_D = 100\%$. At the same time, none of the compared algorithms come close to attaining this P_D for FAR < 0.5.

Remark 3: Given the computational cost of operations involving the Weibull and Rice distributions, such as the modified Bessel function of the first kind, and the good results obtained with simpler distributions, we opted not to perform simulations with these distributions in this article.

The simulation analysis for these distributions depends on approximations and simplifications that will be addressed in future research.

VII. CONCLUSION

Following the paradigm in [33], we propose a CDA based on entropy hypothesis testing for stacks of UWB VHF SAR images. We derive analytical expressions for the entropy and the entropy variance for the Weibull and Rice distributions. To the best of our knowledge, these derivations have not been presented in the (h, ϕ)-entropy literature.

Tests have been conducted over the CARABAS II data for a number of distributions modeling the clutter-plus-noise. We conclude that the best performance in terms of FAR and PD for the intensity radar images analyzed is obtained, in order by the Gaussian, Rayleigh, Gamma, and log-normal models. In addition, we incorporate the ideas of [18] to create a mask based on the median over all stacks of our stack entropy-based statistics. This noticeably reduces the FAR and increases the PD without using any a priori information. Our best scenario, consisting of the Gaussian assumption for the clutter-plus-noise, PB stacks, and median masking, results in less than 0.1 false alarms per square kilometer and perfect detection. This is the only CDA that reaches this level of performance for the CARABAS II dataset.

Future research directions include using other types of stochastic distances for change detection and using asymptotic series, such as the Edgeworth and Gram–Charlier series, to describe probability distributions. New stochastic distances may present computational benefits, while the asymptotic series should provide a more accurate representation of the real statistical nature of the images without much added complexity.

Another avenue of research is to explore the Weibull, Rice, and K distributions more in-depth, obtaining the necessary results to use them to derive entropy-based statistics with feasible time complexity.

APPENDIX
VARIABLES NECESSARY FOR THE COMPUTATION OF THE VARIANCE OF THE ENTROPY OF SELECTED DISTRIBUTIONS

A. Weibull

In what follows, we present the variables a_w, b_w, and c_w, which are used for the computation of the variance of the entropy $\sigma_H^2(\boldsymbol{\theta})$ for the Weibull distribution, where $\boldsymbol{\theta}^{\mathrm{T}} = (\lambda, k)$

$$a_w \triangleq \frac{E_X\left[X^k \ln^2(X)\right]}{\lambda^k} - \frac{2\ln(\lambda)E_X\left[X^k \ln(X)\right]}{\lambda^k}$$
$$+ \ln^2(\lambda) + \frac{1}{k^2} \tag{24}$$

$$b_w \triangleq -\frac{kE_X\left[X^k \ln(X)\right]}{\lambda^{k+1}} \tag{25}$$

$$c_w \triangleq \left(\frac{k}{\lambda}\right)^2 \tag{26}$$

where the expected value is computed according to the Weibull pdf of X in (14).

B. Rice

The variables a_R, b_R, and c_R are used in the computation of the variance of the entropy for the Rice distribution and are defined as

$$a_R \triangleq \frac{1}{\sigma^2} - \frac{2\sigma^2 + \nu^2}{\sigma^4} + \frac{E_X\left[\frac{1}{\nu}X\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right]}{\sigma^2}$$
$$+ \frac{E_X\left[X^2\left(\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right)^2\right]}{\sigma^4}$$

$$b_R \triangleq \frac{4}{\sigma^2} - \frac{2\nu^2}{\sigma^4} - \frac{4\nu^4}{\sigma^6} - \frac{E_X\left[2\nu X\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right]}{\sigma^4}$$
$$+ \frac{E_X\left[4\nu^2 X^2\left(\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right)^2\right]}{\sigma^6}$$

$$c_R \triangleq \frac{2\nu}{\sigma^3} + \frac{2\nu^3}{\sigma^5} - \frac{E_X\left[2\nu X^2\left(\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right)^2\right]}{\sigma^5}.$$

Let $\boldsymbol{\theta}^{\mathrm{T}} = (\nu, \sigma)$. The partial derivatives of the entropy with respect to the distribution parameters ν and σ are also needed for the computation of the variance of the entropy $\sigma_H^2(\boldsymbol{\theta})$ and are given, respectively, by

$$\frac{\partial H}{\partial \nu} = \frac{2\nu}{\sigma^2} + \frac{E_X[\nu \ln(X)]}{\sigma^2}$$
$$- \frac{E_X\left[X \ln(X)\frac{I_1(\nu X/\sigma^2)}{I_0(\nu X/\sigma^2)}\right]}{\sigma^2}$$
$$+ \frac{E_X\left[\nu \ln(I_0(\nu X/\sigma^2)) - X\frac{I_1(\nu X/\sigma^2)}{I_0(\nu X/\sigma^2)}\right]}{\sigma^2}$$
$$- \frac{E_X\left[X \ln(I_0(\nu X/\sigma^2))\frac{I_1(\nu X/\sigma^2)}{I_0(\nu X/\sigma^2)}\right]}{\sigma^2}$$

$$\frac{\partial H}{\partial \sigma} = \frac{2}{\sigma} - \frac{2\nu^2}{\sigma^3} + \left(\frac{2}{\sigma} + \frac{\nu^2}{\sigma^3}\right)E_X[\ln(X)]$$
$$+ \frac{E_X\left[X^2 \ln(X) + 2\nu X \ln(X)\frac{I_1(X\nu/\sigma^2)}{I_0^2(X\nu/\sigma^2)}\right]}{\sigma^3}$$
$$+ E_X\left[\ln(I_0(X\nu/\sigma^2)]\left(\frac{2}{\sigma} + \frac{\nu^2}{\sigma^3}\right)\right.$$
$$+ \frac{E_X\left[X^2 \ln(I_0(X\nu/\sigma^2)) + 2\nu X\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right]}{\sigma^3}$$
$$+ \frac{E_X\left[2\nu X \ln(I_0(X\nu/\sigma^2))\frac{I_1(X\nu/\sigma^2)}{I_0(X\nu/\sigma^2)}\right]}{\sigma^3}.$$

The expected values are computed according to the Rice pdf of an RV X, as in (15).

REFERENCES

[1] A. Asokan and J. Anitha, "Change detection techniques for remote sensing applications: A survey," *Earth Sci. Inform.*, vol. 12, no. 2, pp. 143–160, Jun. 2019.

[2] W. Shi, M. Zhang, R. Zhang, S. Chen, and Z. Zhan, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, p. 1688, May 2020.

[3] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[4] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Volume I—Basic Principles*, 1st ed. London, U.K.: Institution of Engineering and Technology, 2010.

[5] L. Wang et al., "Radial velocity of ocean surface current estimated from SAR Doppler frequency measurements—A case study of Kuroshio in the East China Sea," *Acta Oceanologica Sinica*, vol. 40, no. 12, pp. 135–147, Dec. 2021.

[6] L. Wang et al., "An automated extraction of small- and middle-sized rice fields under complex terrain based on SAR time series: A case study of Chongqing," *Comput. Electron. Agricult.*, vol. 200, Sep. 2022, Art. no. 107232.

[7] L. Zhao, E. Chen, Z. Li, W. Zhang, and Y. Fan, "A new approach for forest height inversion using X-band single-pass InSAR coherence data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5206018.

[8] W. L. Melvin and J. A. Scheer, *Principles of Modern Radar: Radar Applications*, vol. 3. London, U.K.: Institution of Engineering and Technology, Dec. 2013.

[9] G. Smith and L. M. H. Ulander, "A model relating VHF-band backscatter to stem volume of coniferous boreal forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 728–740, Mar. 2000.

[10] M. L. Imhoff, "Radar backscatter and biomass saturation: Ramifications for global biomass inventory," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 511–518, Mar. 1995.

[11] H. Hellsten, L. M. H. Ulande, A. Gustavsson, and B. Larsson, "Development of VHF CARABAS II SAR," *Proc. SPIE*, vol. 2217, pp. 48–60, Apr. 1996.

[12] H. Israelsson et al., "Retrieval of forest stem volume using VHF SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 36–40, Jan. 1997.

[13] M. L. Imhoff, S. Carson, and P. Johnson, "A low-frequency radar experiment for measuring vegetation biomass," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 6, pp. 1988–1991, 1998.

[14] D. I. Alves et al., "A statistical analysis for wavelength-resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 227–231, Feb. 2020.

[15] R. Machado, V. T. Vu, M. I. Pettersson, P. Dammert, and H. Hellsten, "The stability of UWB low-frequency SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1114–1118, Aug. 2016.

[16] L. M. H. Ulander, W. E. Pierson, M. Lundberg, P. Follo, P.-O. Frolind, and A. Gustavsson, "Performance of VHF-band SAR change detection for wide-area surveillance of concealed ground targets," *Proc. SPIE*, vol. 5427, pp. 259–270, Sep. 2004.

[17] L. M. H. Ulander, M. Lundberg, W. Pierson, and A. Gustavsson, "Change detection for low-frequency SAR ground surveillance," *IEE Proc.-Radar, Sonar Navigat.*, vol. 152, no. 6, pp. 413–420, Dec. 2005.

[18] B. G. Palm et al., "Wavelength-resolution SAR ground scene prediction based on image stack," *Sensors*, vol. 20, no. 7, pp. 1–16, Apr. 2020.

[19] V. T. Vu, "Wavelength-resolution SAR incoherent change detection based on image stack," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1012–1016, Jul. 2017.

[20] D. I. Alves et al., "Wavelength-resolution SAR change detection using Bayes' theorem," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5560–5568, 2020.

[21] D. I. Alves et al., "Neyman–Pearson criterion-based change detection methods for wavelength-resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[22] H. Hellsten and R. Machado, "Bayesian change analysis for finding vehicle size targets in VHF foliage penetration SAR data," in *Proc. IEEE Radar Conf.*, Oct. 2015, pp. 510–515.

[23] H. Hellsten, R. Machado, M. I. Pettersson, V. T. Vu, and P. Dammert, "Experimental results on change detection based on Bayes probability theorem," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 318–321.

[24] J. Neyman and E. S. Pearson, "IX. On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London. Ser. A, Math. Phys. Eng. Sci.*, vol. 231, pp. 289–337, Feb. 1933.

[25] V. T. Vu, N. R. Gomes, M. I. Pettersson, P. Dammert, and H. Hellsten, "Bivariate gamma distribution for wavelength-resolution SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 473–481, Jan. 2019.

[26] V. T. Vu, M. I. Pettersson, R. Machado, P. Dammert, and H. Hellsten, "False alarm reduction in wavelength-resolution SAR change detection using adaptive noise canceler," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 591–599, Jan. 2017.

[27] L. P. Ramos et al., "A wavelength-resolution SAR change detection method based on image stack through robust principal component analysis," *Remote Sens.*, vol. 13, no. 5, p. 833, Feb. 2021.

[28] J. G. Vinholi, D. Silva, R. Machado, and M. I. Pettersson, "CNN-based change detection algorithm for wavelength-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[29] J. G. Vinholi, B. G. Palm, D. Silva, R. Machado, and M. I. Pettersson, "Change detection based on convolutional neural networks using stacks of wavelength-resolution synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5236414.

[30] A. B. Campos, M. I. Pettersson, V. T. Vu, and R. Machado, "False alarm reduction in wavelength-resolution SAR change detection schemes by using a convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[31] M. Salicru, M. L. Menendez, D. Morales, and L. Pardo, "Asymptotic distribution of (h, φ)-entropies," *Commun. Statist.-Theory Methods*, vol. 22, no. 7, pp. 2015–2031, Jan. 1993, doi: 10.1080/03610929308831131.

[32] L. Pardo, D. Morales, M. Salicrú, and M. L. Menéndez, "Large sample behavior of entropy measures when parameters are estimated," *Commun. Statist.-Theory Methods*, vol. 26, no. 2, pp. 483–501, Jan. 1997.

[33] A. C. Frery, R. J. Cintra, and A. D. C. Nascimento, "Entropy-based statistical analysis of PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3733–3743, Jun. 2013.

[34] M. Lundberg, L. M. H. Ulander, W. E. Pierson, and A. Gustavsson, "A challenge problem for detection of targets in foliage," *Proc. SPIE*, vol. 6237, pp. 160–171, May 2006.

[35] X. Qin, S. Zhou, H. Zou, and G. Gao, "A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 806–810, Jul. 2013.

[36] G. Gao, S. Gao, K. Ouyang, J. He, and G. Li, "Scheme for characterizing clutter statistics in SAR amplitude images by combining two parametric models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5636–5646, Oct. 2018.

[37] G. H. Mittmann Voigt et al., "A statistical analysis for intensity wavelength-resolution SAR difference images," *Remote Sens.*, vol. 15, no. 9, p. 2401, May 2023.

[38] V. T. Vu, M. I. Pettersson, and N. R. Gomes, "Stability in SAR change detection results using bivariate Rayleigh distribution for statistical hypothesis test," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 37–40.

[39] N. R. Gomes, P. Dammert, M. I. Pettersson, V. T. Vu, and H. Hellsten, "Comparison of the Rayleigh and K-distributions for application in incoherent change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 756–760, May 2019.

[40] D. I. Alves et al., "Change detection method for wavelength-resolution SAR images based on bayes' theorem: An iterative approach," *IEEE Access*, vol. 11, pp. 84734–84743, 2023.

[41] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Raleigh, NC, USA: SciTech Publishing, 2004.

[42] G. Gao, "Statistical modeling of SAR images: A survey," *Sensors*, vol. 10, no. 1, pp. 775–795, Jan. 2010.

[43] V. Anastassopoulos and G. A. Lampropoulos, "High resolution radar clutter classification," in *Proc. Int. Radar Conf.*, 1995, pp. 662–667.

[44] T. Musha and M. Sekine, *Advanced Radar Techniques and Systems*. London, U.K.: Institution of Engineering and Technology, 1993, pp. 153–183.

[45] H. Ogawa, M. Sekine, T. Musha, M. Aota, and M. Ohi, "Weibull-distributed radar clutter reflected from sea ice," *IEICE Trans.*, vol. 70, pp. 116–120, Feb. 1987.

[46] M. Sekine, T. Musha, H. Ogawa, and M. Chikara, "Effect of pulse length changes on Weibull clutter of sea ice," *IEE Proc. F Radar Signal Process.*, vol. 137, no. 1, pp. 15–18, 1990.

[47] M. Sekine et al., "Weibull-distributed ground clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 596–598, Jul. 1981.

[48] J. B. Billingsley, A. Farina, F. Gini, M. V. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 2, pp. 579–593, Apr. 1999.

[49] S. Sayama and H. Sekine, "Weibull, log-weibull and K-distributed ground clutter modeling analyzed by AIC," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 3, pp. 1108–1113, Jul. 2001.

[50] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.