Multi-Model CFAR Detection in FOliage PENetrating SAR Images

Alessio Izzo*, Marco Liguori*, Carmine Clemente†,
Carmela Galdi*, Maurizio Di Bisceglie* and John J. Soraghan†

* Department of Engineering, Università degli Studi del Sannio, 82100 Benevento, Italy
E-mail: alessiodiesel@hotmail.it, marco_494@hotmail.it, galdi@unisannio.it, dibisceg@unisannio.it

† University of Strathclyde, CeSIP, EEE, 204, George Street, G1 1XW, Glasgow, UK
E-mail: carmine.clemente@strath.ac.uk, j.soraghan@strath.ac.uk

Abstract

A multi-model approach for Constant False Alarm Ratio (CFAR) detection of vehicles through foliage in FOliage PENetrating (FOPEN) SAR images is presented. Extreme value distributions and Location Scale properties are exploited to derive an adaptive CFAR approach that is able to cope with different forest densities. Performance analysis on real data is carried out to estimate the detection and false alarm probabilities in the presence of a ground truth.

I. Introduction

The fundamental feature of FOPEN Radars is the capability to collect returns from scatterers under foliage [1]. This is achieved by using fairly low carrier frequencies (in the UHF and VHF bands) that are able to penetrate the vegetating layer. The ability to ”see” through foliage canopies makes FOPEN radar a powerful tool for military purposes and, more interestingly, if SAR techniques are used, a FOPEN SAR sensor becomes able to detect, track and recognize vehicles hiding in forests [1].

Due to the nature of the imaged scene, several issues are still under investigation for the complete and reliable exploitation of such sensors. In particular, canopies and hidden vehicles are not the only possible reflecting targets in a forest scene; trunks are present and contribute significantly to the intensity of the signal returned to the radar. Detections will be affected from reflections from trunks if an accurate strategy of control of false alarms is not provided. Solutions or partial solutions to this problem have been provided in literature. Clutter modelling has been identified as a viable solution to mitigate tree and trunks detections, physical, statistical and the combination of the two approaches were used to model forest clutter in FOPEN SAR [2], [3], [4], [5], [6]. The models proposed in [2], [3], [4] consider electromagnetic modeling of forests to extract deterministic clutter models. These models are useful in terms of understanding of the scattering physics but are not applicable in statistical detection frameworks. The model presented in [5], introduces statistical properties in the model in [3], however this model is not robust with respect to the presence of tree trunks dominating a scene. In [6] a model for VHF clutter generation was proposed, integrating

both background scatterers and large-amplitude discrete clutter (trees). Despite its flexibility, the model proposed in [6] does not have a closed form solution and therefore it is not suitable to derive detectors with false alarm rate control. Different detection approaches have been evaluated in literature; in [7] a two parameters CFAR, a change detection and multi aperture detection approaches have been evaluated, with the conclusion that a standard CFAR approach does not perform as well as the other two approaches. In [8] a rank-order filters based approach was used to address the poor performance of a traditional CFAR detector, finally the case of slow moving target detection was investigated in [9] where the authors proposed an alternative approach based on Along Track Interferometry.

The main problem in VHF SAR clutter is that the data samples are highly inhomogeneous and variable. In contrast to corner reflector effects at microwaves frequencies, targets do not exhibit the high peak response in FOPEN SAR images.

Due to the lack of accurate analytic models for FOPEN SAR clutter, a robust algorithm is required against large variability of amplitude and distribution model of the received data.

Starting from a statistical modeling of the clutter, in this paper we introduce a novel framework based on multi-model CFAR detection in adverse clutter conditions exhibited from FOPEN SAR images.

The Multi Model approach derives by a generalization of the single model. In selecting a specific model for target detection the designer is assuming a risk; the risk of model mismatch is mitigated thanks to the flexibility of our approach that can be used for any type of LS distributions. The de-risk is obtained through the exploitation of a goodness of fit based clutter model selection. [10].

In our approach, the clutter is statistically modeled by light and heavy-tailed distributions that belongs (directly or after transformations) to the location-scale (LS) family. The light and heavy-tailed distributions are consistent with the nature of the scattering from different forest scenarios [1], [6], whereas the location-scale family is a requirement of the CFAR detection algorithm [11] that is exploited in this work.

The CFAR detection algorithm introduced in [12], [13] and applied to high resolution SAR images in [11] can be embedded in a framework that is able to select the best location-scale distribution to be used to compute the adaptive threshold and that will ensure the Constant False Alarm Rate in the highly inhomogeneous FOPEN SAR image environment. Generally, it is not easy to prove that a Goodness of fit test based CFAR detector has the CFAR property, but when the data distribution has some special properties we can prove that the resultant detector is CFAR. One of these cases is when the assumed distribution is a member of location-scale (LS) family.

The paper is organized as follows, Section II introduces the multi-Model CFAR detection framework, addressing the specific cases of Gumbel for Maximum and Weibull distributed background. The performance in terms of distribution fitting of the two above mentioned

models are assessed and discussed in Section III using real VHF FOPEN SAR data. Section IV discusses the algorithm performance on real data, demonstrating the capability of the proposed approach to control the false alarm probability and to detect extended targets hiding in foliage.

NOTATION

Throughout the paper, we use (lower case) boldface for vectors $\mathbf{x}$, and (upper case) for matrices $\mathbf{X}$. The transpose and inverse matrix operators are denoted by the symbol $(\cdot)^T$ and $(\cdot)^{-1}$, respectively. After sorting and censoring operations, the vector $\mathbf{x}_i$ is denoted as $\tilde{\mathbf{x}}_i$; instead $\tilde{\mathbf{x}}_i'$ represent the vector $\tilde{\mathbf{x}}_i$ after the transformation function $f_i(\cdot)$ related by i-th distribution.

II. MULTI-MODEL CFAR DETECTOR IN LS ENVIRONMENT

The architecture of the proposed algorithm is depicted in Figure 1.

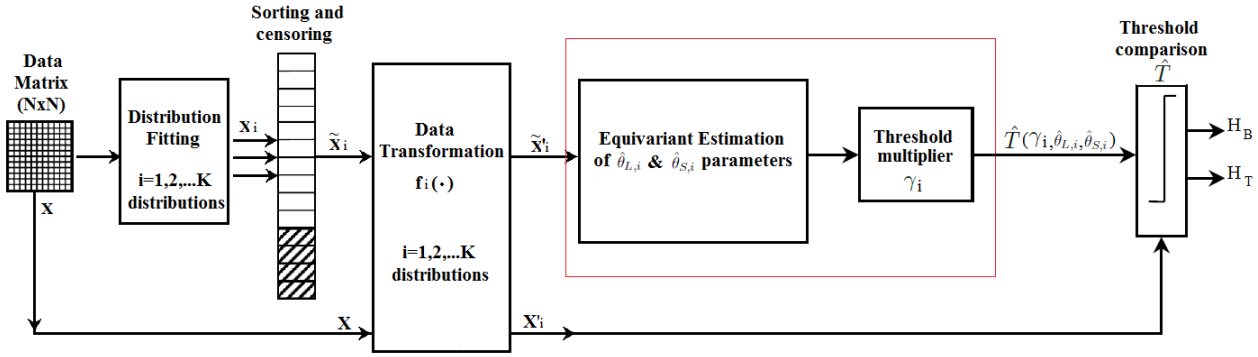


Figure 1: Architecture of the Multi-Model CFAR system in LS environment.

The algorithm has been designed in order to ensure higher robustness and reliability of the results, with respect to the single model approach introduced in [11], by considering K possible statistical distributions for the background. The multi-Model approach stems as a generalization of the single model approach where, by selecting a specific distribution model, it is assumed a risk of model mismatch. This risk is mitigated thanks to the flexibility of the approach that is applicable to any family of LS distributions[1].

[1]A random variable X is of the Location Scale type, with *location* parameter $\theta_L \in \mathbb{R}$ and *scale* parameter $\theta_S > 0$, if any variate of the family can be obtained by an affine transformation of the standardized variate X_0, namely

$$X = \theta_S X_0 + \theta_L$$

where it is clearly understood that the standardized variate corresponds to $\theta_S = 1$ and $\theta_L = 0$. As a consequence, the CDF has the property:

$$F_X(x; \theta_L, \theta_S) = F_{X_0}\left(\frac{x - \theta_L}{\theta_S}\right)$$

where $F_{X_0(\cdot)}$ is the distribution of the standardized variate X_0 [11].

In FOPEN SAR, the multi-Model detector automatically searches for the distribution that best fits the real data in a specific reference window, introducing robustness with respect to non accurate a-priori knowledge of trees density in a given cell.
The problem can be posed in terms of the binary hypothesis test:

$$\begin{cases} H_{B|D_i}: & X(m,n) \sim F_i(x) & \text{Target absent} \\ H_{T|D_i}: & X(m,n) \sim F_{T_i}(x) & \text{Target present} \end{cases} \tag{1}$$

where (m,n) represents the cell under test. In the $H_{B|D_i}$ hypothesis the probability distribution of the background sample is one of a selected set $\{F_i(x)\}_{i=1\ldots K}$ of distributions with cardinality K. The probability distribution is selected among the K distributions of Location-Scale (LS) type using a minimum distance non-parametric test. In the hypothesis $H_{T|D_i}$ the presence of a target in the background is modeled by the distribution $F_{T_i}(x)$, that is unknown. Afterwards, a CFAR procedure is applied for the selected background hypothesis.
With reference to Figure 1 it is assumed that the following statements are applicable.

- The data window for the estimation of the background statistics is composed of $N \times N$ intensity samples obtained from the complex SAR image.
- The statistical characterization of the background is chosen according to a best-fit strategy between the Empirical Cumulative Distribution Function (ECDF) of the data and the CDFs from a set of model distributions. The fitting strategy is based on the Lilliefors test, a modified Kolmogorov-Smirnov test accounting for the case of unknown distributional parameters [14].

  The usual outcome of the Lilliefors test, with a given significance level α, is:

  - H_0: the selected distribution is compatible with data, i.e. the p-value is greater than α;
  - H_1: the null hypothesis is rejected, i.e. the p-value is less than α.

  However, we are here interested in fitting the best distribution model rather than determining if the H_0 hypothesis holds given a significance level α. Therefore, we use the $p-$value to determine the a-posteriori significance level for the distribution model that best fits the data.
- The background is highly non homogeneous due to the presence of tree trunks and foliage [6] and statistical distributions with light or heavy tail characteristics are preferred in such case. To our knowledge, confirmed by simulation assessments, best performance in terms of goodness of fit are provided by:

  - the Gumbel for Maximum CDF

$$F(\mathbf{x}; \theta_L, \theta_S) = exp\left[-exp\left(-\frac{x-\theta_L}{\theta_S}\right)\right] \quad \text{with} \quad \theta_L \in \mathbb{R},\ \theta_S > 0 \tag{2}$$

    which belongs to a location-scale family;

- the Weibull CDF

$$F\left(\mathbf{x};\kappa,\lambda\right)=\begin{cases}1-e^{-\left(\frac{x}{\lambda}\right)^{k}}, & \text{if } x\geq 0,\\ 0, & \text{if } x<0.\end{cases} \tag{3}$$

that is a location-scale (Gumbel) distribution after a logarithmic transformation. Location and scale parameters are derived from scale and shape parameters of the Weibull CDF as in [11], [12].

Let ρ_i be $p-$value associated to the $i-$th distribution $F_i(x)$, for each reference window the selection rule for the statistical distribution is:

$$F_i(x) : i=\underset{i=1,2,\ldots,K}{\arg\max}\,\rho_i \tag{4}$$

- Data are organized into a vector $\mathbf{x}_i$ that is sorted and censored. After censoring, that consists in removing the largest r samples in the vector $\mathbf{x}_i$, the new vector $\tilde{\mathbf{x}}_i$ is produced. This transformation is required in order to control the self masking from targets inside the reference window [15].
- The Multi-Model CFAR algorithm is applied by exploiting, for each reference window, the parameters $(\gamma_i,\hat{\theta}_{L,i},\hat{\theta}_{S,i})$ of the specific distribution, after a proper transformation to get an LS distribution. For the case at hand the data transformation follows the rule

$$f_i\left(\cdot\right)=\begin{cases}1 & \text{for } i=1\\ ln\left(\cdot\right) & \text{for } i=2\end{cases} \tag{5}$$

The Best Linear Unbiased (BLU) estimates of the Location and Scale parameters are obtained by minimizing the variance of the estimators subject to the constraints of unbiasedness [13]. It is found that

$$\begin{pmatrix}\hat{\theta}_{L,i}\\ \hat{\theta}_{S,i}\end{pmatrix}=\left(\mathbf{H}_i^T\mathbf{C}_{0,i}^{-1}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^T\mathbf{C}_{0,i}^{-1}\tilde{\mathbf{x}}_{\mathbf{i}}' \tag{6}$$

where $\tilde{\mathbf{x}}_{\mathbf{i}}'$ is a vector of ordered and censored samples from a LS distribution, $\mathbf{H}_i=(\mathbf{1}\,\boldsymbol{\mu}_{0,i})$ and $\boldsymbol{\mu}_{0,i}$ and $\mathbf{C}_{0,i}$ are the mean vector and the covariance matrix of the standardized vector $\tilde{\mathbf{x}}_{0,i}'$, respectively, and whose expressions are reported in Appendix A. Location and scale parameters related to the specific distribution under test can be thus estimated through (6), and the threshold for CFAR detection can be derived as

$$\hat{T}\left(\gamma_i,\hat{\theta}_{L,i},\hat{\theta}_{S,i}\right)=\hat{\theta}_{S,i}(\tilde{\mathbf{x}}_{\mathbf{i}}')\,\gamma_i+\hat{\theta}_{L,i}(\tilde{\mathbf{x}}_{\mathbf{i}}'). \tag{7}$$

The quantities in (11), in Appendix A, can be used to evaluate, for any value of the censoring depth r, the mean vector and the covariance matrix of the censored data. In other words, the "censored" mean vector as well as the "censored" covariance matrix can be obtained from the "uncensored" expressions by discarding the last r elements and the last r rows and columns, respectively. However, the evaluation of the adaptive

threshold in (7) requires also the knowledge of the *threshold multiplier* γ_i, that is evaluated according to the desired false alarm probability.
The threshold multiplier is the solution of the equation:

$$P_{FA}|D_i = Pr\left\{ \frac{X_i'(m,n) - \hat{\theta}_{L,i}}{\hat{\theta}_{S,i}} > \gamma_i |\, H_{B|D_i} \right\}, \tag{8}$$

which is the $(1 - P_{FA})$-quantile of the normalized test statistic $[(X_i'(m,n) - \hat{\theta}_{L,i})/\hat{\theta}_{S,i}]$, where $X_i'(m,n)$ is the sample under test. Obviously, if the statistical distribution of the test statistic is known, then γ_i can be determined. Unfortunately, this distribution cannot be evaluated in a closed form because it requires the knowledge of the joint distribution of the variable $X_i'(m,n)$ and of the location and scale estimators. More easily, the value of the quantile γ_i can be computed via Monte Carlo simulation where N_p realizations of the test statistic are generated and the threshold multiplier level is estimated from the empirical CDF [16]. In order to improve the estimators reliability, a suitable number of trials is needed.

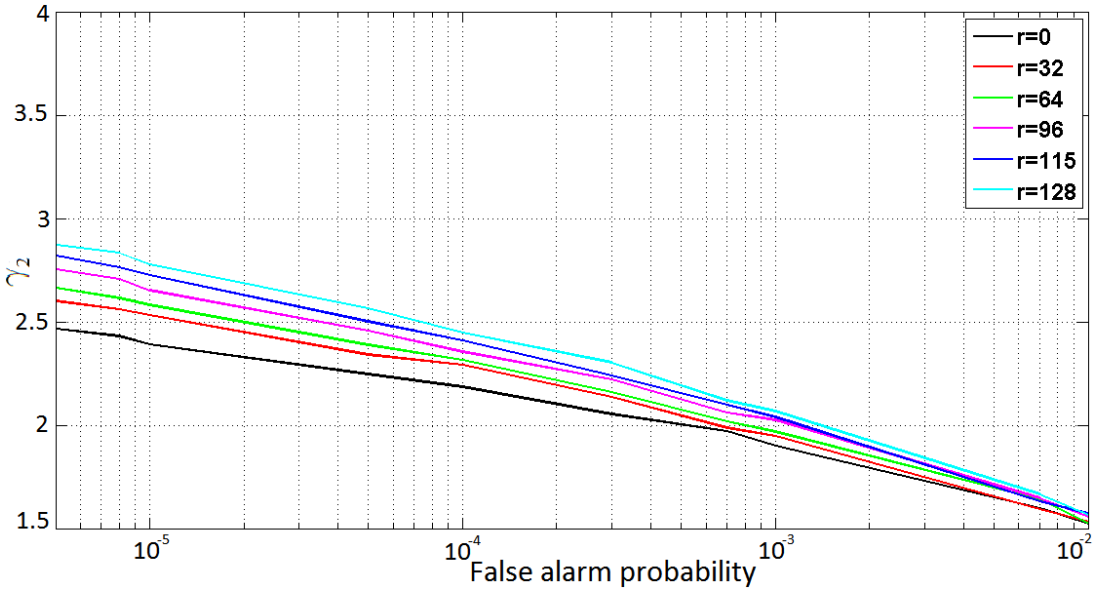


Figure 2: Threshold multiplier values versus P_{FA}, for a Log-Weibull distribution, with $M = N \times N = 256$ and various censoring depth r.

In Figure 2 plots show the value of the threshold multiplier versus the false alarm probability for a Log-Weibull distribution with $N = 16$, $M = 256$ and various values of the censoring depth r. As seen in Figure 2 the value of threshold multiplier increases by reducing the false alarm probability.

- Finally, the decision rule

$$X_i'(m,n) \underset{H_{B|D_i}}{\overset{H_{T|D_i}}{\gtrless}} \hat{T}\left(\gamma_i, \hat{\theta}_{L,i}, \hat{\theta}_{S,i}\right) \tag{9}$$

  compares each data pixel with the adaptive threshold for CFAR detection. The decision rule (9) is applied using the threshold value $\hat{T}\left(\gamma_i, \hat{\theta}_{L,i}, \hat{\theta}_{S,i}\right)$ associated to the specific distribution.

The algorithm adapts itself with respect to the distribution that fits better the real data in a specific reference window, with no a-priori knowledge of trees density.
As can be seen from equation 8, the false alarm probability is conditionally CFAR. Nevertheless, if the same false alarm probability is kept for all single models, the multi model approach keeps the CFAR property [2].

Proof. *Considering* $P_{FA}|D_i = \tilde{P}_{FA}$, $\forall i = 1 \dots K$, *for the law of total probability*

$$P_{FA} = \sum_{i=1}^{K} Pr\left(D_i\right) \cdot P_{FA}|D_i = \tilde{P}_{FA} \cdot \sum_{i=1}^{K} Pr\left(D_i\right) = \tilde{P}_{FA} \tag{10}$$

where $\sum_{i=1}^{K} Pr\left(D_i\right) = 1$, *and* P_{FA} *is the total false alarm probability of the multi model algorithm.* □

III. Statistical characterization of background

In this section we present the results obtained in terms of goodness of fit for the selection of the statistical distributions in the H_B hypothesis. The dataset was acquired using the Swedish low frequency SAR system *CARABAS-II VHF SAR* [17]. The system transmits HH-polarized radio waves between 20-90 MHz, corresponding to wavelengths between 3.3 m and 15 m. In the imaged areas 25 military vehicles are concealed by forest, in four deployments (for reader's convenience see [17]). Due to the presence of trees in the scenario, we deal with extremely inhomogeneous data. During setup, several light and heavy tailed distributions were analysed, e.g. the Rayleigh, Normal, Log-Normal, Gamma, Weibull and Extreme Value Distributions. The Log-Normal distribution in low density forest is rejected in the 42.33% of the analysed cells. Weibull Distribution and the Gumbel for Maximum Distribution provide the best results for low and high density forests, respectively. For this reason, and for shortness, these two distributions have been selected and results will be discussed in this section.

[2]CFAR property ensured an estimated P_{FA} values in agree with the designed value when the correct model is chosen during the statistical characterization of the background stage.

A. *High Density Forest*

In order to analyse the goodness of fit of the data to the distributions under test, we consider the area shown in Figure 4-a. Within this image, we selected the subimage enclosed in the red frame in Figure 4-b, that contains an homogeneous area without targets, with high density of trees. The selected area has the same characteristics of the surrounding area, where the targets are placed. For this area, a set of 29×29 reference windows, indexed through the respective quadrant in the Figure 4-b, represents the dataset for the statistical validation. Each reference window is composed of 16×16 pixels producing a set of 256 samples. For each reference window, a Lilliefors test is carried out to verify the compatibility of the data with the design distribution. To this aim, the empirical CDF has been compared to the theoretical CDF, considering the Gumbel for Maximum distribution for high density forest data. This comparison is shown in Figure 3.

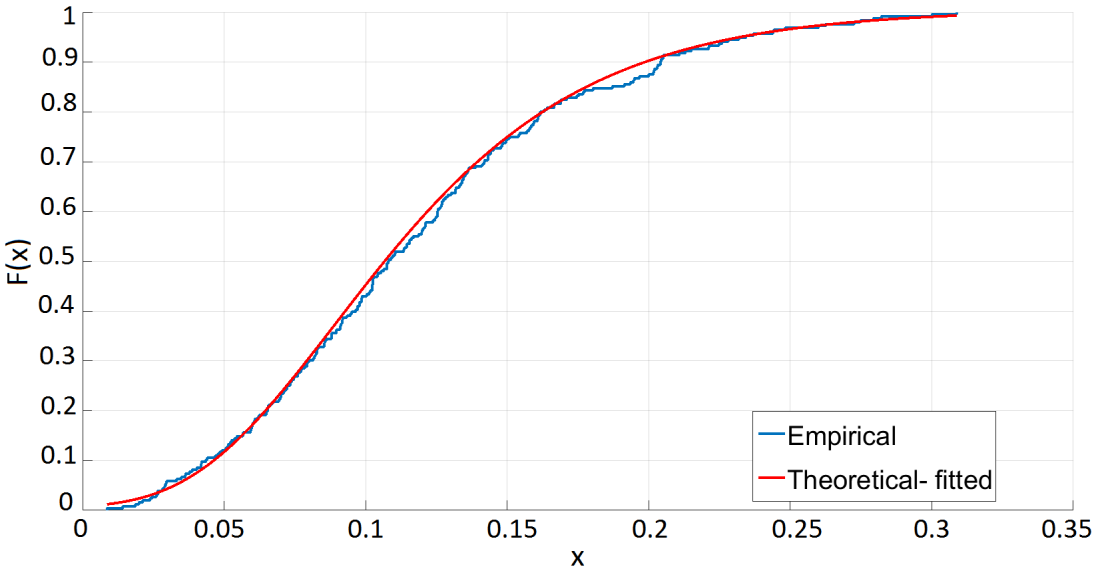


Figure 3: Empirical CDF compared to the theoretical Gumbel for Maximum CDF for a reference window of 16×16 pixels, in a high density forest.

The p-values for each reference window in the homogeneous area of Figure 4-b, are evaluated for both H_0 hypothesis, where the Gumbel for Maximum distribution fits well the real data, and H_1 hypothesis, where the Gumbel for Maximum distribution is rejected (with a significance level set to 0.1%).

Varying the reference window size, as shown in the Table I, best results are obtained with a reference window of 16×16 pixels. In this case, it is found that H_0 hypothesis was rejected 26 times, corresponding to a percentage of acceptance of the Gumbel for Maximum distribution of 96.91%.

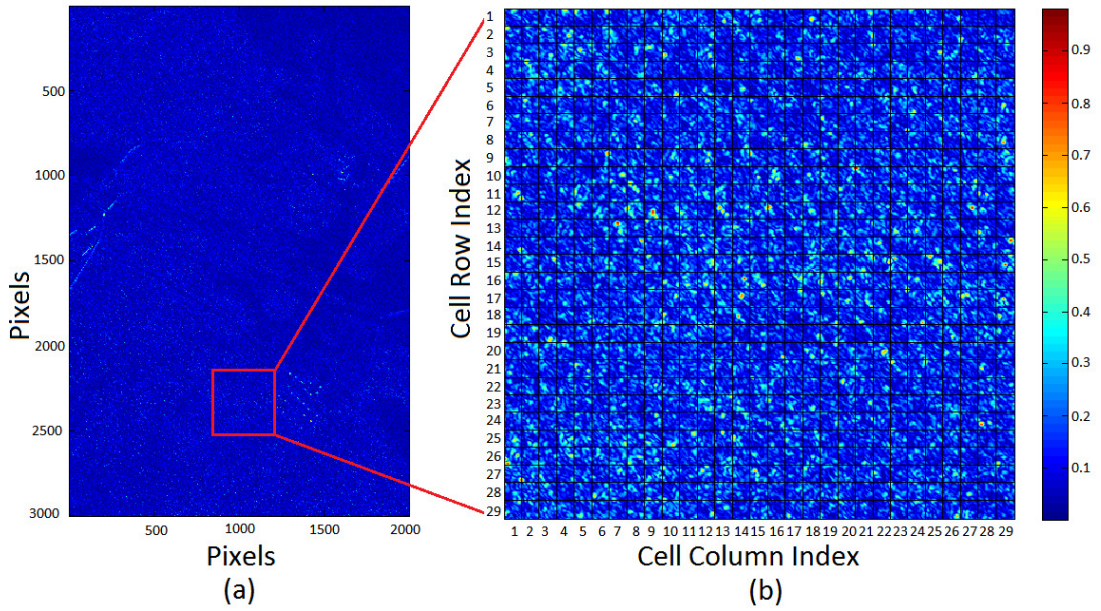


Figure 4: a) SAR image of 2000×3000 samples, related to flight pass number 1 of Fredrik mission. b) The homogeneous area under test with pixel coordinates [2101-2565,701-1165], within forest 1 [17].

B. Low Density Forests

In a low density forest, the homogeneous area into Forest 2 dataset [17] is considered. The statistical characterization was carried out by considering an image of the flight pass 1 of Sigismund mission. In Figure 5 a set of 29×29 reference windows is used for test, with reference window size equal to 16×16 pixel.

For low density forest data the Weibull distribution is considered. For each reference window, a Lilliefors test is carried out to verify the compatibility of the data with the design distribution. The comparison between the empirical CDF and the theoretical Weibull is shown in Figure 6.

From Table I, we note that the Weibull Distribution shows a good fit in 97.50% of cases (H_0 hypothesis was rejected 21 times). By changing the reference window dimensions a slight performance degradation was found for low density forest scenario as already seen in high density forest case; 16×16 reference window size, provides the best fitting, for both Weibull and Gumbel for Maximum distributions.

C. Multi model fitting

Best performance are achieved through the implementation of a multi model fitting approach as described in Section II. In order to evaluate the advantage of this approach in both low and high density forests, the same homogeneous areas of Forest 1 in Figure 4 and Forest 2 in Figure 5 were considered.

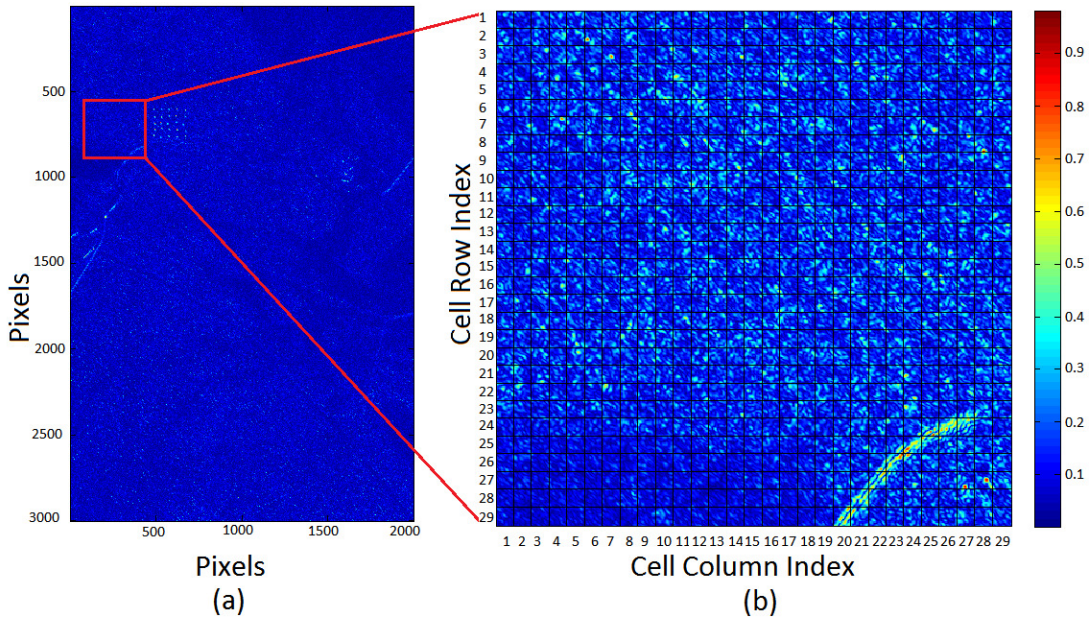


Figure 5: a) SAR image of flight pass number 1 of Sigismund mission, showing 2000 $\times$ 3000 acquired samples. b) The homogeneous area under test with coordinates [451-915,1-465] in Forest 2.

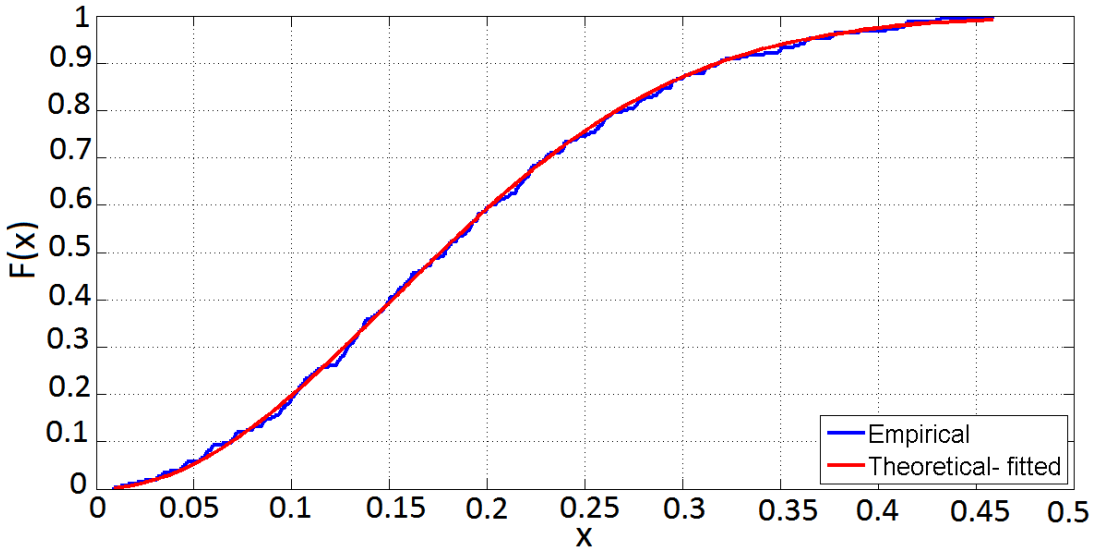


Figure 6: Empirical CDF compared to the theoretical Weibull CDF for a reference window of 16 $\times$ 16 pixels, in a low density forest.

Differently from the single model approach, whereby either the Gumbel or Weibull distributions have been exploited, with the multi model approach the Lilliefors test is applied considering both the distributions and choosing the one that provides the highest p-value, greater than the significance level. The H_0 percentage of Lilliefors test outcomes obtained

Table I: Lilliefors test outcomes for high density and low density forests: percentages where the H_0 hypothesis using Gumbel Maximum distribution and Weibull distribution is accepted.

Fredrik Mission	**Gumbel Maximum Distribution**
16×16	$\mathbf{96.91}\%$
24×24	$\mathbf{93.35}\%$
32×32	$\mathbf{89.80}\%$
Sigismund Mission	**Weibull Distribution**
16×16	$\mathbf{97.50}\%$
24×24	$\mathbf{94.46}\%$
32×32	$\mathbf{92.86}\%$

using the multi model fitting approach, for homogeneous area of Figure 4 and 5, increases in both scenarios.

For high density forest, from a 96.91% of H_0 acceptance percentage, associated to the Gumbel for Maximum single model, a 98.57% has been achieved using the multi model algorithm.

For low density forest, by comparing Weibull single model approach with the multi model approach, we find that the percentage of H_0 hypothesis is increased from 97.50%, in the Weibull single model, to 99.16% in the multi model approach (H_0 is rejected only 7 times in the second case).

It is worth to remark that the H_0 hypothesis is rejected only for the sake of statistical validation, but in the detection scheme a distribution is always selected according to the maximum p-value, even if it is lower than the significance level.

IV. Performance analysis of the CFAR detector

In order to demonstrate the CFAR property of the single and multi-model approaches, simulated data have been considered. In particular, a $P_{FA} = 10^{-4}$ was set like designed false alarm rate and $100/P_{FA}$ realizations of simulated data composed by N^2 samples are drawn exactly from both Weibull and Gumbel distributions. Successively, sorting and censoring block is applied, accounting for the number of valid samples available for each censoring depth r.

The P_{FA} probability are reported in the tables II and III, choosing the locations θ_L and scale θ_S parameters from a uniform distribution $\mathcal{U}(a, b)$ with a and b values tuned on the real data. In particular for the Gumbel case $(a_l = 0.051, b_l = 0.275)$ and $(a_s = 0.029, b_s = 0.157)$ are considered for the location and scale parameters respectively, while $(a_s = 0.770, b_s = 0.397)$ and $(a_c = 1.450, b_c = 2.824)$ are used for the scale and shape parameter of the Weibull distribution. For the single model, how clearly reported in the first column of the tables II and III, the P_{FA} values oscillate around the designed P_{FA} ensuring the CFAR property.

In addiction, a Model Miss-Match Analysis has been done. In this case, simulated data from a chosen distribution has been draught, using a single-model CFAR detector based on the other distribution. How clearly reported in the second column of the tables II and III, the P_{FA}

Table II: Simulated data from Gumbel distribution: false alarm probabilities using both the correct single-model, the wrong single-model and multi-model CFAR detector for various values of the censoring depth, with a designed $P_{FA} = 10^{-4}$.

Censoring Depth	P_{FA}_GUM	P_{FA}_WBL	P_{FA}_MM
0	1.19×10^{-4}	1.3×10^{-3}	2.21×10^{-4}
32	1.11×10^{-4}	2.22×10^{-3}	2.48×10^{-4}
64	9.28×10^{-5}	2.29×10^{-3}	2.49×10^{-4}
96	1.04×10^{-4}	3.30×10^{-3}	2.67×10^{-4}
115	9.49×10^{-5}	3.13×10^{-3}	2.61×10^{-4}
128	1.01×10^{-4}	2.90×10^{-3}	2.68×10^{-4}

Table III: Simulated data from Weibull distribution: false alarm probabilities using both the correct single-model, the wrong single-model and multi-model CFAR detector for various values of the censoring depth, with a designed $P_{FA} = 10^{-4}$.

Censoring Depth	P_{FA}_WBL	P_{FA}_GUM	P_{FA}_MM
0	9.88×10^{-5}	3.9×10^{-6}	7.67×10^{-5}
32	8.86×10^{-5}	3.72×10^{-6}	6.15×10^{-5}
64	9.98×10^{-5}	3.38×10^{-6}	5.83×10^{-5}
96	1.08×10^{-4}	5.71×10^{-6}	5.38×10^{-5}
115	1.02×10^{-4}	7.55×10^{-6}	4.69×10^{-5}
128	1.04×10^{-4}	1.01×10^{-5}	4.81×10^{-5}

values don't respect the designed P_{FA}. The same procedure has been done with the multi model approach. The outcomes reported in the tables II and III, confirm the effectiveness of the proposed approach ensuring robustness of mixed scenario. The discrepancies between the single and multi-model results are due mainly to cases in which the Lilliefors test is not able to discriminate correctly the true data distribution. Furthermore, in this section we investigate the performance of the single-model and multi-model CFAR detector for different forest environments, using real data. The CFAR detector performance are optimized by choosing the best size for the reference window and for the depth of censoring. A reasonable rule is to take the number of samples in the reference window much greater than the maximum expected object size and, at the same time, to discard a number of samples that is at least equal to the overall size of the objects. The efficiency of the algorithm has been tested for different values of the reference window size, censoring depth and false alarm rate. The attention will focus on the flight pass number 1 of each mission/target deployment [17] with the following parameters: P_{FA}=10^{-4}; $N = 16$; $r = [0 \quad 32 \quad 64 \quad 96 \quad 115 \quad 128]$. Single-model, model mismatch and multi-model analysis for high and low density forests are considered and the results are discussed in the next sections.

A. False Alarm Rate

The first analysis is performed for an area with high density within Forest 1 scenario, in absence of targets (see Figure 4). The single-model CFAR detector is applied using a Gumbel for maximum distribution and the false alarm probabilities for single and multi-model CFAR

detector are reported in Table IV, for a design false alarm rate equal to 10^{-4} and for various values of the censoring depth. The detectors exhibit P_{FA} values that are compatible with the design rate, however the multi-model approach ensures higher reliability thanks to its capability to adapt to the statistical model.

Table IV: Real data from Fredrik deployment: false alarm probabilities using both the correct single-model (Gumbel Max), the wrong single-model (Weibull) and multi-model CFAR detector for various values of the censoring depth, with a designed $P_{FA} = 10^{-4}$..

Censoring Depth $\mathbf{r}$	P_{FA}_GUM	P_{FA}_WBL	P_{FA}_MM
0	5.55×10^{-5}	2.54×10^{-4}	7.86×10^{-5}
32	0.97×10^{-4}	1.11×10^{-3}	1.76×10^{-4}
64	1.20×10^{-4}	1.17×10^{-3}	1.80×10^{-4}
96	1.66×10^{-4}	2.00×10^{-3}	2.45×10^{-4}
115	1.99×10^{-4}	2.01×10^{-3}	3.05×10^{-4}
128	2.40×10^{-4}	2.01×10^{-3}	3.56×10^{-4}

A similar analysis is considered for the Forest 2 dataset (low forest density, in absence of targets) shown in Figure 5. Both the single-model CFAR detector for low density forest (using a Weibull distribution) and the multi-model approach have been considered. The resulting false alarm probabilities are reported in Table V, for a design false alarm rate equal to 10^{-4} and various values of the censoring depth. Again, the false alarm probabilities are compatible with the design value.

Table V: Real data from Sigismund deployment: false alarm probabilities using both the correct single-model (Weibull), the wrong single-model (Gumbel Max) and multi-model CFAR detector for various values of the censoring depth, with a designed $P_{FA} = 10^{-4}$.

Censoring Depth $\mathbf{r}$	P_{FA}_WBL	P_{FA}_GUM	P_{FA}_MM
0	1.11×10^{-4}	0	4.16×10^{-05}
32	3.00×10^{-4}	1.37×10^{-5}	6.94×10^{-05}
64	4.53×10^{-4}	1.38×10^{-5}	4.42×10^{-05}
96	5.87×10^{-4}	1.39×10^{-5}	4.16×10^{-05}
115	4.81×10^{-4}	1.39×10^{-5}	5.55×10^{-05}
128	5.97×10^{-4}	1.39×10^{-5}	7.86×10^{-05}

B. Detection Probability

The detection capability of the single and multi- model CFAR detector is first evaluated by considering the area of Forest 1. Figure 7-a and -b shows the SAR image of 5×5 array of targets and the targets ground truth [3].

The single-model CFAR detector has been applied using the design values $P_{FA} = 10^{-4}$ and various censoring depths. In Table VI the probability of detection for a single - model CFAR with Gumbel for Maximum detector with P_{FA} set to 10^{-4} are reported along with the

[3] It has to been considered that the targets ground truth has been estimated affecting the detection probability performance.

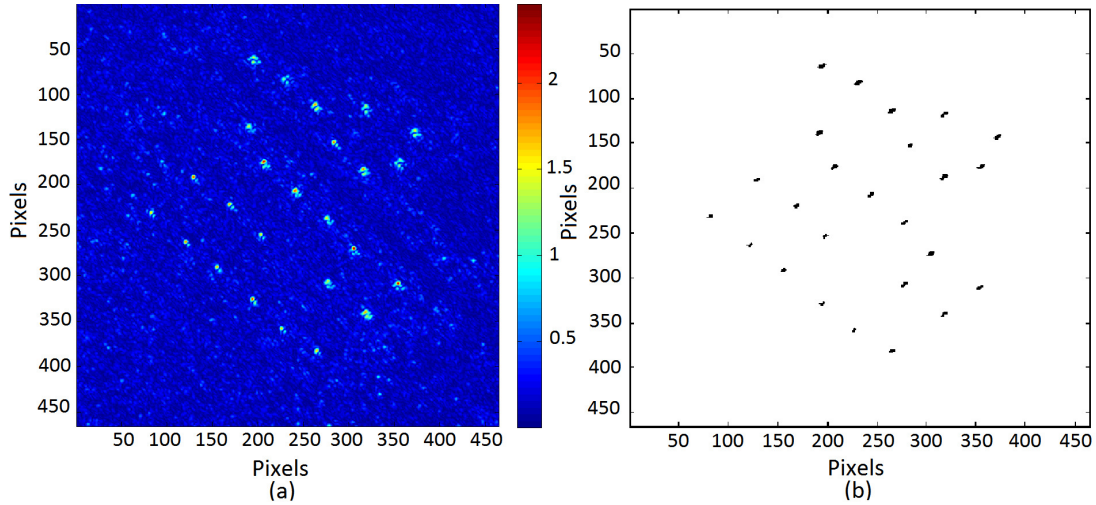


Figure 7: a) SAR image of the 5 by 5 targets array within forest 1 characteristic of Fredrik deployment; b) Targets ground truth for Fredrik deployment.

probabilities of detection of a multi-model CFAR detector referred to the same area under test.

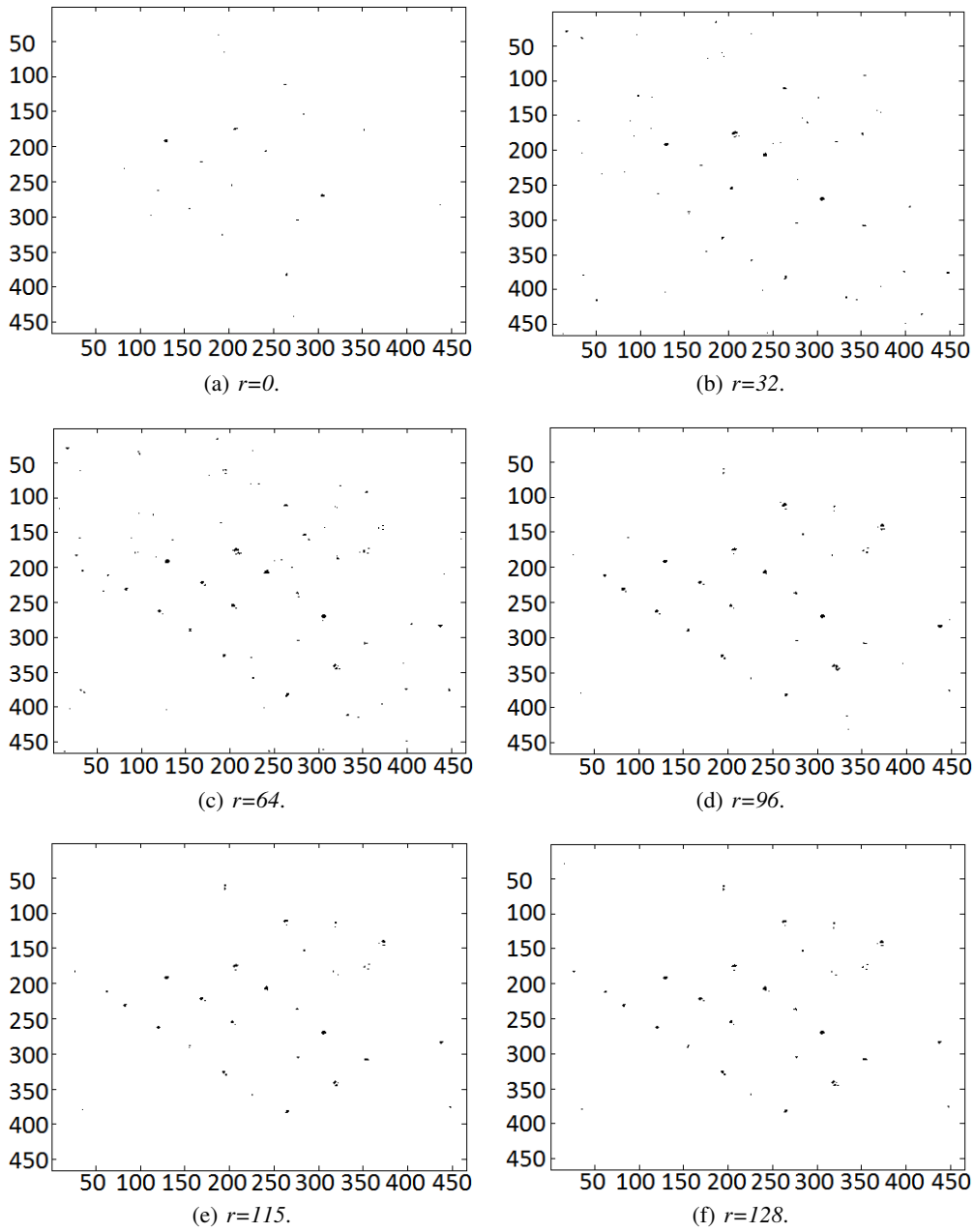
When no censoring is applied the targets are not correctly detected, due to the self-masking effect from the targets. More correct detections are obtained when r is increased up to 128 samples. Further increase of the depth of censoring does not introduce additional features to the targets detection but, otherwise, generates underestimation of the distribution parameters.

Table VI: Detection probabilities for a single-model CFAR Gumbel for maximum detector vs multi-model CFAR detector in Fredrik deployment.

Censoring Depth **r**	P_D_MM	P_D_GUM
0	0.076	0.076
32	0.125	0.124
64	0.172	0.170
96	0.211	0.209
115	0.222	0.220
128	0.237	0.233

Detection maps for Multi-Model CFAR detector are reported in Figure 8, for various censoring depths and $P_{FA} = 10^{-4}$. Comparing Figure 8-c with 8-f, where an increase of the depth of censoring is applied, we can clearly note the performance optimization in terms of false alarms.

The Multi-Model CFAR algorithm achieves equivalent performance in terms of detection probabilities as the Single-Model CFAR algorithm (see Table VI). The most important remark is the higher accuracy of the multi-model approach in the estimate of the background parameters, hence a higher reliability of the obtained detections. To complete the performance analysis, the low-density area in Forest 2 shown in Figure 9-a is considered. The targets empirical



(a) *r=0*. (b) *r=32*.

(c) *r=64*. (d) *r=96*.

(e) *r=115*. (f) *r=128*.

Figure 8: Detection maps for multi-model CFAR detector, referred to Fredrik deployment (Forest 1) and varying censoring depths; $P_{FA} = 10^{-4}$ and cell size 16×16.

ground truth in Figure 9-b, has been considered to quantify the capability of the algorithm to detect targets under foliage.

In Table VII the probability of detection for a single - model CFAR Weibull detector with P_{FA} set to 10^{-4} is reported along with the probability of detection of a multi-model CFAR detector referred to the same area under test. The detection maps for the multi-model detection

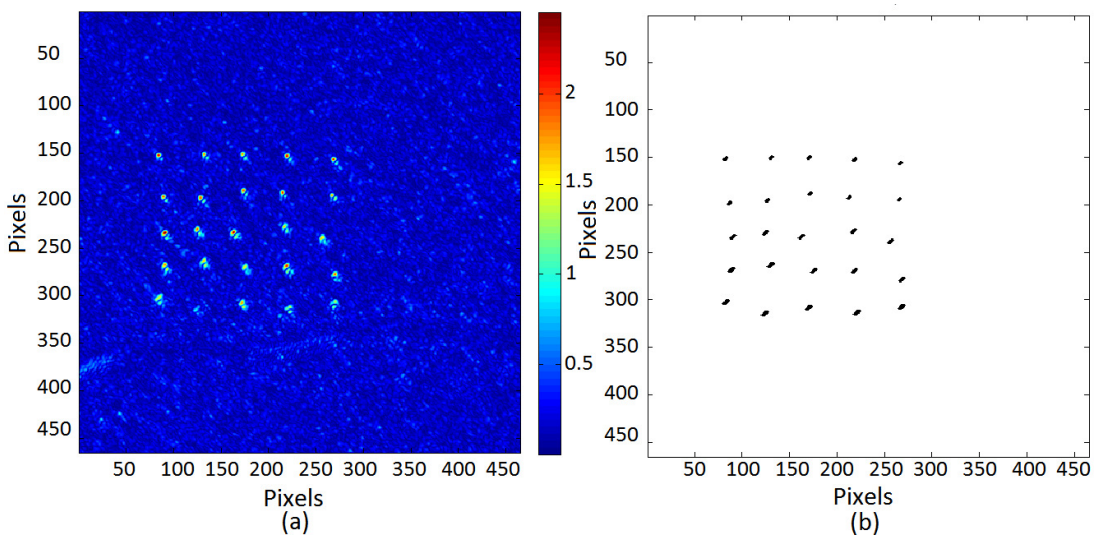


Figure 9: a) SAR image representing the 5 by 5 targets array within forest 2 characteristic of Sigismund deployment; b) Targets ground truth for Sigismund deployment.

Table VII: Detection probabilities for a single-model CFAR Weibull detector vs multi-model CFAR detector for Sigismund deployment.

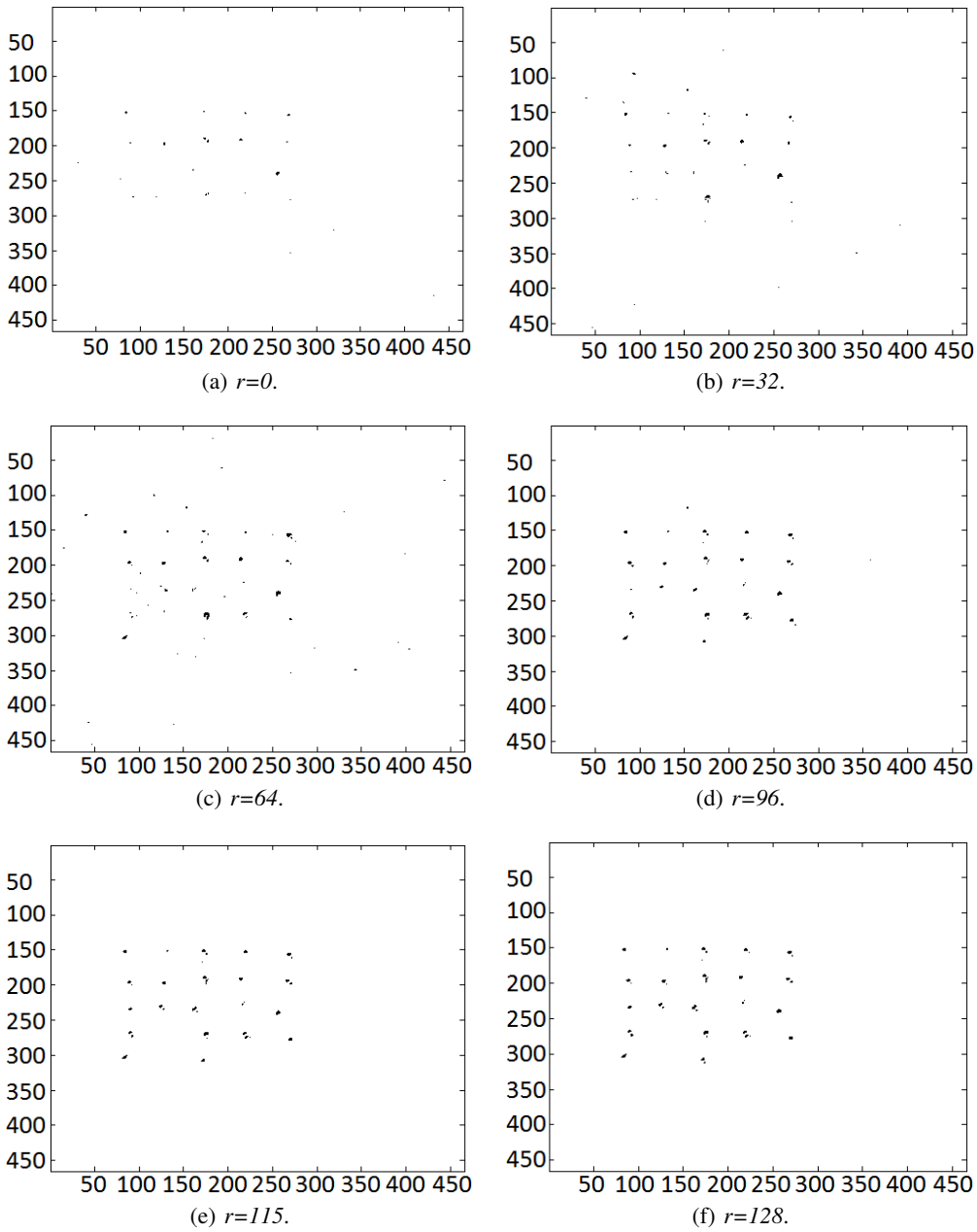
Censoring Depth **r**	P_D_MM	P_D_WBL
0	0.105	0.045
32	0.167	0.120
64	0.257	0.277
96	0.355	0.396
115	0.373	0.434
128	0.394	0.468

cases reported in Table VII are shown in Figure 10.

C. Clustering strategy

We now assess the capability to detect extended targets using the proposed algorithm. In particular, by increasing the design false alarm probability, a higher detection probability is achievable. By introducing some domain knowledge like the size of the targets of interest, an a-posteriori analysis can be performed to decrease the false alarm rate (preserving the high detection probability). Precisely, by using a **clustering strategy**[4] and target size selection, the extended targets can be extracted from the detection map. The clustering algorithm classifies the detections according to a pixels aggregation procedure and assigns a value to every cluster equal to the number of pixels forming it. Then, using a-priori knowledge about targets size, clusters smaller than a fixed threshold can be discarded. This operation leads to a strong decrement of the false alarms. Setting the design P_{FA} of a Multi-Model CFAR algorithm to

[4]Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups [18]. In this work the "aggregation" method is exploited, an algorithm based on the aggregation of adjoining pixels.



(a) *r=0*. (b) *r=32*.

(c) *r=64*. (d) *r=96*.

(e) *r=115*. (f) *r=128*.

Figure 10: Detection maps for multi-model CFAR detector, referred to Sigismund deployment (Forest 2) and varying censoring depths; $P_{FA} = 10^{-4}$ and cell size 16×16.

10^{-2} an higher number of detected pixels is achieved. An area with targets within forest 1 with Fredrik deployment (as shown in Figure 7) is considered. The clustering algorithm has been applied to the area under test. Following the clustering the distribution of the detected target sizes can be obtained, as shown in Figure 11. Exploiting the a-priori domain knowledge about the size of the smaller target ($\approx 14m^2$), the clusters smaller than that value have been

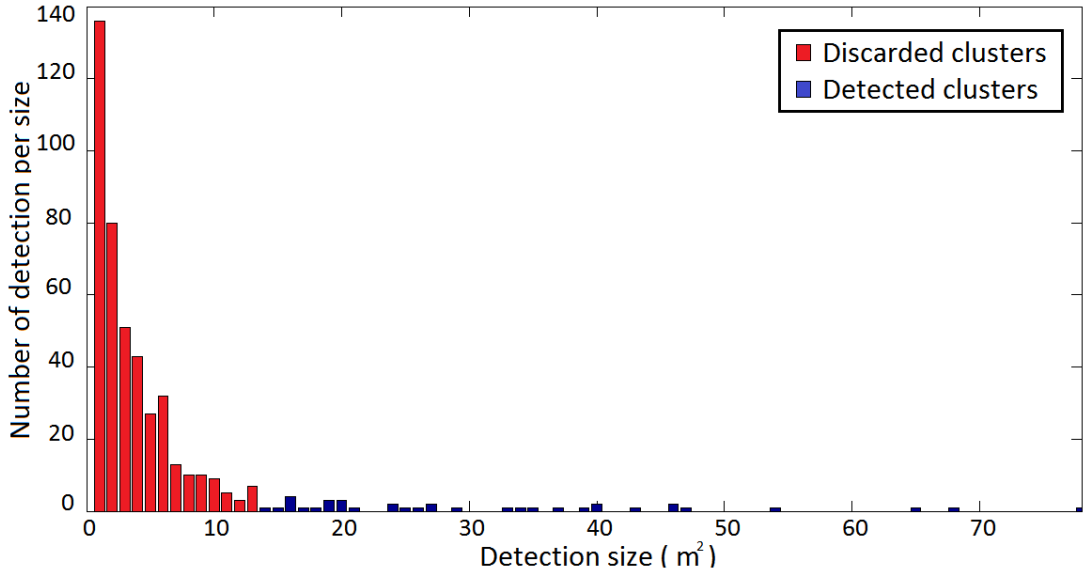


Figure 11: Histogram of number of detections per size vs detection size. In red are shown the target with size less than 14 m^2 while in blue are shown the targets with size greater or equal than 14 m^2.

discarded (red bars in Figure 11). The resulting detection map after the clustering and the size threshold is shown in Figure 12, where the red detections can be discarded.

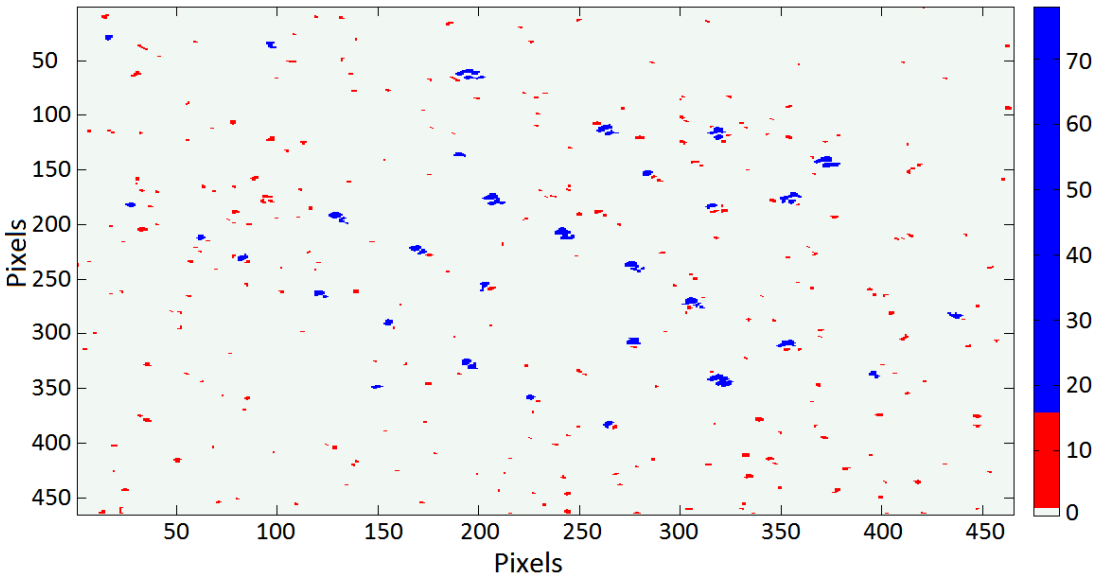


Figure 12: Clustering strategy: Detection map of a Multi-Model CFAR detector with $P_{FA} = 10^{-2}$ and $r = 128$. The red color represents the discarded pixels, instead the blue color highlights the potential targets.

In Table VIII the performance in terms of false alarms and detection probabilities on a pixel

Table VIII: Multi-model CFAR detector vs the same detector with Clustering strategy: Performance analysis.

$P_{FA} = 10^{-2}$, $r = 128$	P_D	P_{FA}
Multi-Model CFAR	0.5804	1×10^{-2}
Multi-Model CFAR with Clustering	0.5804	3×10^{-3}

basis before and after the clustering and target size threshold are reported. In this case the proposed approach provides a bounded False Alarm Rate. The clustering strategy applied to the Multi-Model CFAR algorithm provides a decrement of the false alarm probability, keeping the same detection probability. Hence, the clustering algorithm can be used if an higher capability of detecting extended targets is required. For example, for the case in hand, the 96% of targets were detected (1 vehicle only was missed).

V. Conclusion

In this paper a novel framework for CFAR detection of extended target in FOPEN SAR images has been proposed using a multi-model approach. The novel framework exploits a CFAR detection algorithm based on location-scale distributions. Due to the peculiar problem in hand, the location-scale property was not the only constraint on the selection of the statistical distribution to be considered but, due to the presence of trees in the scene, distributions having light and heavy-tails were considered as well. The proposed framework is able to control the False Alarm Probability in the FOPEN SAR scenario, adapting the best location-scale distribution fitting the background data. The effectiveness of the proposed approach has been demonstrated on real data, showing that the framework is able to adapt both the model and the threshold, and provides a higher level of reliability. Moreover, the capability to detect extended targets has been demonstrated by exploiting domain knowledge. The usefulness of the Multi Model approach, derived by generalization of the single model, is reinforced by the possibility of using it also for other kind of clutter, since that it could be used for any types of LS distribution. In this paper, it has been used for FOPEN clutter, but it does not exclude the possibility of using this approach for SAR data collected in other conditions. For example sea clutter at low grazing angle exhibits sea spikes that can lead to a behaviour of the radar returns similar to those of large trunks in forests. Future work will deal with the generalization of the statistical model and to the application to 3D FOPEN SAR.

Acknowledgment

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing and the University of Sannio. The dataset of Carabas II SAR images

can be obtained at the link: https://www.sdms.afrl.af.mil/index.php?collection=vhf_change_detection.

APPENDIX A

Letting $G_i^{(m)}$ be the m-th order derivative of the inverse cumulative distribution function $G_i(\cdot) = F_i^{-1}(\cdot)$, the approximate expressions of mean and covariance matrix of the ordered samples are [11]:

$$\begin{aligned}
\mu_{0(k),i} &\approx G_{k,i} + \frac{p_k q_k}{2(n+2)} G_{k,i}^{(2)} + \frac{p_k q_k}{(n+2)^2} \\
&\quad \times \left(\frac{1}{3}(q_k - p_k) G_{k,i}^{(3)} + \frac{1}{8} p_k q_k G_{k,i}^{(4)} \right), \\
C_{0(k,h),i} &\approx \frac{p_k q_h}{(n+2)} G_{k,i}^{(1)} G_{h,i}^{(1)} + \frac{p_k q_h}{(n+2)^2} \\
&\quad \times [(q_k - p_k) G_{k,i}^{(2)} G_{h,i}^{(1)} + (q_h - p_h) G_{h,i}^{(2)} G_{k,i}^{(1)} \\
&\quad + \frac{1}{2} p_k q_k G_{k,i}^{(3)} G_{h,i}^{(1)} + \frac{1}{2} p_h q_h G_{k,i}^{(1)} G_{h,i}^{(3)} \\
&\quad + \frac{1}{2} p_k q_h G_{k,i}^{(2)} G_{h,i}^{(2)}], \qquad k \leq h
\end{aligned} \tag{11}$$

with $p_k = k/(n+1)$ and $q_k = (1 - p_k)$ and $G_{k,i} = G(p_k)$. The moments in (11) can be evaluated with reference to a specific reduced distribution by direct substitution of the derivatives (up to the fourth order) of its quantile function (inverse CDF). Let $G_{i=1}(x)$ be the quantile function associated with the reduced extreme value distribution of type I for maximum (Gumbel for Maximum), the set of equations to substitute in (11), and containing the four derivatives under test are:

$$\begin{aligned}
G_1(x) &= -ln\left(-ln(x)\right); \\
G_1^{(1)}(x) &= -\left(\frac{1}{x\, ln(x)} \right); \\
G_1^{(2)}(x) &= \left(\frac{ln(x)+1}{x^2\, ln^2(x)} \right) \\
G_1^{(3)}(x) &= -\left(\frac{2\, ln^2(x) + 3\, ln(x) + 2}{x^3\, ln^3(x)} \right) \\
G_1^{(4)}(x) &= \left(\frac{6\, ln^3(x) + 11\, ln^2(x) + 12\, ln(x) + 6}{x^4\, ln^4(x)} \right).
\end{aligned} \tag{12}$$

The same procedure can be applied to obtain the quantile function associated to the reduced extreme value distribution of type I for minimum (Log-Weibull).

In this case, we obtain:

$$
\begin{aligned}
&G_2(x) = ln\left(-ln(1-x)\right);\\
&G_2^{(1)}(x) = \left(\frac{1}{(x-1)\,ln(1-x)}\right);\\
&G_2^{(2)}(x) = -\left(\frac{ln(1-x)+1}{(x-1)^2\,ln^2(1-x)}\right)\\
&G_2^{(3)}(x) = \left(\frac{2\,ln^2(1-x)+3\,ln(1-x)+2}{(x-1)^3\,ln^3(1-x)}\right)\\
&G_2^{(4)}(x) =\\
&-\left(\frac{6\,ln^3(1-x)+11\,ln^2(1-x)+12\,ln(1-x)+6}{(x-1)^4\,ln^4(1-x)}\right).
\end{aligned}
\tag{13}
$$

REFERENCES

[1] Davis, M. E., *Foliage Penetration Radar. Detection and Characterization of Objects Under Trees*, Scitech Publishing, 2012.

[2] Fransson, J., Walter, F., Ulander, L., "Estimation of forest parameters using CARABAS-II VHF SAR data," *IEEE Transactions on Geoscience and Remote Sensing,*, vol. 38, no. 2, pp. 720–727, Mar 2000.

[3] Smith, G. Ulander, L., "A model relating VHF-band backscatter to stem volume of coniferous boreal forest," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 728–740, Mar 2000.

[4] Williams, M., Manninen, T., Kellomaki, S., Ikonen, V.-P., Sievanen, R., Lehtonen, M., Nikinmaa, E., Vesala, T., "Modeling the SAR response of pine forest in Southern Finland," in *Proceedings. 2003 IEEE International Geoscience and Remote Sensing Symposium, 2003. IGARSS '03.*, July 2003, vol. 2, pp. 1350–1352 vol.2.

[5] Kononov, A. Ka, M.-H., "Model-Associated Forest Parameter Retrieval Using VHF SAR Data at the Individual Tree Level," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 69–84, Jan 2008.

[6] Jackson, J. A. Moses, R. L., "A Model for Generating Synthetic VHF SAR Forest Clutter Images," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1138 – 1152, 7 2009.

[7] Nanis, J., Halversen, S., Owirka, G., Novak, L., "Adaptive filters for detection of targets in foliage," *IEEE Aerospace and Electronic Systems Magazine*, vol. 10, no. 8, pp. 34–36, Aug 1995.

[8] Mitra, A., Lewis, T., Shaw, A., "Rank-order filters for FOPEN target detection," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 93–96, Feb 2004.

[9] Kapfer, R. Davis, M., "Along track interferometry for foliage penetration Moving Target Indication," in *IEEE Radar Conference, 2008. RADAR '08.*, May 2008, pp. 1–6.

[10] Norouzi, Y., Gini, F., Nayebi, M., Greco, M., "Non-coherent radar CFAR detection based on goodness-of-fit tests," *IET Radar, Sonar and Navigation*, vol. 1, no. 2, pp. 98–105, April 2007.

[11] Bisceglie, M. D. Galdi, C., "CFAR Detection of Extended Objects in High-Resolution SAR Images ," *IEEE Transaction on Geoscience and Remote Sensing*, 2005.

[12] Longo, M., Guida, M., Lops, M., "Biparametric linear estimation for CFAR against Weibull clutter," 1992.

[13] Conte, E., Lops, M., Tulino, A., "Hybrid procedure for cfar in non-gaussian clutter," *IEE Proceedings Radar, Sonar and Navigation*, vol. 144, no. 6, pp. 361–369, Dec 1997.

[14] Conover, W., *Practical nonparametric statistics*, Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1980.

[15] Richards, M., Scheer, J., Scheer, J., Holm, W., *Principles of Modern Radar*, Number v. 1 in Principles of Modern Radar. SciTech Publishing, Incorporated, 2010.

[16] Kay, S. M., *Fundamentals of Statistical Signal Processing volume II: Detection Theory*, Prentice-Hall, 1998.

[17] Hellsten, H., Ulander, L., Gustavsson, A., Larsson, B., *Development of VHF CARABAS II SAR*, Proc. SPIE vol. 2747, Radar Sensor Technology, held in Orlando, FL, 8-9 April 1996, pag. 48-60, 1996.

[18] Kaufman, L. Rousseeuw, P., *Finding Groups in Data: an Introduction to Cluster Analysis*, Wiley and Sons, 1990.