Computer-Simulated UWB VHF SAR Targets Based on Statistics of Real Data

Bruna G. Palm, Dimas I. Alves, Renato Machado, Viet T. Vu, and Mats I. Pettersson

***Abstract*— In this paper, we performed a statistical study of the targets of wavelength-resolution synthetic aperture radar (SAR) images and addressed the problem of limited training samples. Based on suitable statistical models for the real targets of the evaluated SAR image, we considered a sample generator method to create computer-simulated targets. For such, well-known statistical models, namely, Burr, Gamma, Gaussian, Log-normal, Rayleigh, Rician, and Weibull, are considered for the Anderson-Darling (AD) goodness-of-fit test. This test is a statistical non-parametric hypothesis test generally used to investigate whether a given probability distribution yields a good fit for a given data set. The images used in this study were obtained from the Swedish ultrawideband (UWB) very-high-frequency (VHF) SAR CARABAS II system whose consists of 24 magnitude single-look SAR images. The obtained computer-simulated targets followed similar statistical characteristics when compared to the real ones.**

***Keywords*— CARABAS II, Computer-simulated targets, SAR images, Statistical analysis**

I. Introduction

The identification and classification of distinct ground types [1], [2], modeling [3], and change detection [4], [5] are classical tasks in the context of synthetic aperture radar (SAR) statistical image processing. In particular, wavelength-resolution low-frequency SAR systems are usually used for natural disasters monitoring, foliage-penetrating applications, and detection of concealed targets [6].

Change detection algorithms (CDAs) have been widely considered over the years for the above SAR-related challenges [7], [8]. In particular, the wavelength-resolution SAR change detection is an important topic of research and has been studied for more than a decade [9]. Wavelength resolution SAR systems have also shown unique results with high detection probability associated with a low rate of false alarms per square km, as shown, for example, in [8], [10]. Generally, a CDA is employed to identify changes in a ground scene between distinct measurements in time, such as human-made interference or natural disasters like floods and wildfires [6]. In particular, in wavelength-resolution systems, a CDA can be simply obtained by computing the difference between two single-look images (reference and surveillance), followed by a thresholding operation.

In summary, a CDA classifies the image pixels in changed and unchanged pixels based on previous information extracted from the evaluated images. Consequently, a limited number of training samples may introduce bias on the CDA's result. To solve this problem, we employed a sample generation method—here referred to as 'computer-simulated target', to create samples from the imaging procedure perspective. For such, we considered a statistical study of the real targets. Indeed, the simulation process for a limited number of training samples is an interest research topic for image classification and target detection, as shown, for example, in [11]–[14].

Thus, in this paper, our goal is twofold. First, we performed a statistical study of wavelength-resolution SAR image targets. We considered the 24 wavelength-resolution SAR images provided in [15]. As the pixel magnitudes generally follow an asymmetric distribution and assume nonnegative values, well-known statistical models with such characteristics, namely, Burr, Gamma, Gaussian, Log-normal, Rayleigh, Rician, and Weibull are considered for the Anderson-Darling (AD) test [16], which is the most powerful test based on empirical distribution functions [17]. Second, to address the problem of limited training samples, a sample generator method is utilized to create computer-simulated targets. This process is based on the suitable statistical models for the real targets of the evaluated SAR image. It is expected that the computer-simulated targets follow similar statistical characteristics when compared to the real ones.

The remainder of this paper is organized as follows. Section II describes the database considered in this study. Section III provides a statistical analysis of the image targets, showing the numerical results of the AD test and the obtained computer-simulated targets. Finally, Section IV presents some concluding remarks and directions for future studies.

II. Data Description

The public data set used in this study was obtained from the Swedish ultrawideband (UWB) very-high-frequency (VHF) SAR CARABAS II system, and the images are available in [15]. The data consist of 24 magnitude single-look SAR images that were calibrated, pre-processed, and geocoded. The data set can be divided into three stacks with eight images each, i.e., two out of six passes have identical flight headings. Each image is represented as a matrix of $3{,}000 \times 2{,}000$ pixels, and all images cover the same ground area of 6 square kilometers ($2\,\text{km} \times 3\,\text{km}$).

The ground scene is dominated by boreal forest with pine trees. Fences, power lines, and roads are also present in

B. G. Palm is with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos - SP, Brazil, e-mail: brunagpalm@gmail.com; Dimas I. Alves is with Federal University of PAMPA, Brazil and Federal University of Santa Catarina, Brazil, e-mail: dimasalves@unipampa.edu.br; R. Machado is with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos - SP, Brazil, e-mail: rmachado@ita.br; Viet T. Vu and Mats I. Pettersson are with Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Sweden, e-mail: viet.thuy.vu,mats.pettersson@bth.se. This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Edital Pró-Defesa IV, Brazil.

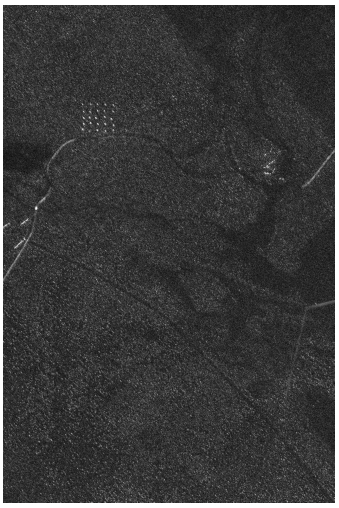


Fig. 1. Image sample from CARABAS II data set. In this image, the vehicles are deployed in the top left of the scene.

the scene. The military vehicles are deployed in the SAR scene and placed uniformly in a manner to facilitate their identifications in the tests [9]. Each image has 25 targets (white dots) with three different sizes, including (i) ten small vehicles with a square design, with dimensions $4.4 \times 1.9 \times 2.2$ m—here defined as 'small size'; (ii) eight truck-sized vehicles with dimensions $6.8 \times 2.5 \times 3.0$ m—here defined as 'medium size'; and (iii) seven trucks with dimensions $7.8 \times 2.5 \times 3.0$ m—here defined as 'large size' [9]. For illustration, one image sample from the CARABAS II data set is shown in Figure 1. In this particular image, the vehicles are obscured by foliage and deployed in the top left of the scene.

III. Statistical Analysis

As mentioned earlier, each image has 25 targets with three different sizes. To identify the statistical characteristics of the targets, we extracted them from the images according to their size in three data sets. Then, we examined several candidates, i.e., distributions for each data set using the AD test, aiming to identify the most suitable distribution for each target size.

For the context of SAR images, the pixel magnitudes usually follow an asymmetric distribution and have nonnegative values; hence, aiming at having a good characterization of image pixel magnitude, we selected some well-known distributions with nonnegative support, such as Burr, Gamma, Log-normal, Rayleigh, Rician, and Weibull. Additionally, we considered the Gaussian distribution as a preliminary study to fit the targets related to the small size data, since this distribution is widely used in signal and image processing [18], [19].

A. Anderson–Darling Goodness-of-Fit Test

To perform the statistical test analysis, we considered the AD test, which is a statistical nonparametric hypothesis test generally used to investigate whether a given probability distribution null hypothesis $\mathcal{H}_0$ yields a good fit for a given sample data [20], [21]. The AD goodness-of-fit test measures the distance A_n between the empirical distribution function $F_Y(y)$ and the cumulative distribution function under the null hypothesis $F_0(y)$ as [22]:

$$A_n = n \int_{-\infty}^{\infty} \left(F_Y(y) - F_0(y)\right)^2 w(y) dF_0(y), \qquad (1)$$

considering the ordered sample values $y[1], y[2], \ldots, y[n]$, where $w(y)$ is a weight function and n is the number of evaluated data points. In particular, the weight function for the AD test is defined as:

$$w(y) = \left[F_0(y)\left(1 - F_0(y)\right)\right]^{-1}.$$

The AD test is computed comparing the distance A_n with a critical value τ. When $A_n > \tau$, the AD test rejects the null hypothesis. The critical value is selected as [22]:

$$\Pr\left(A_n \geq \tau \mid \mathcal{H}_0\right) = 1 - \Pr\left(A_n < \tau \mid \mathcal{H}_0\right) = \alpha,$$

where α is the significance level.

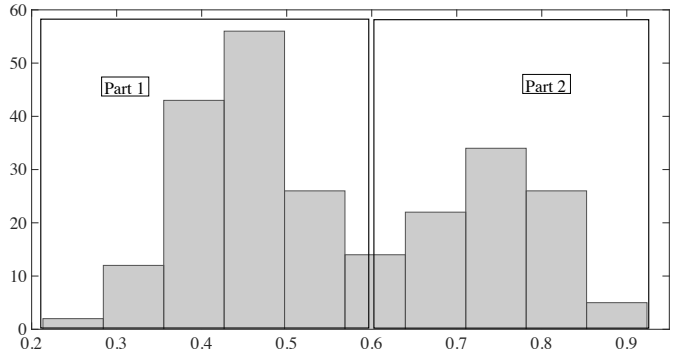
B. Numerical Results

This section presents the numerical results of the statistical analysis retrieved from the AD test for the seven tested distributions: Burr, Gamma, Gaussian, Log-normal, Rayleigh, Rician, and Weibull. To perform the AD test, we set $\alpha = 0.05$, which is a typical significance level value considered for this kind of test [23]. All parameters are estimated from the target magnitudes using the maximum likelihood estimator. We considered three window sizes: 5×5, 7×7, and 9×9 pixels around the central pixel of each target to examine the statistical models. For such, we computed the pixel window average. These window sizes guarantee that the three target sizes considered in this study may be covered.

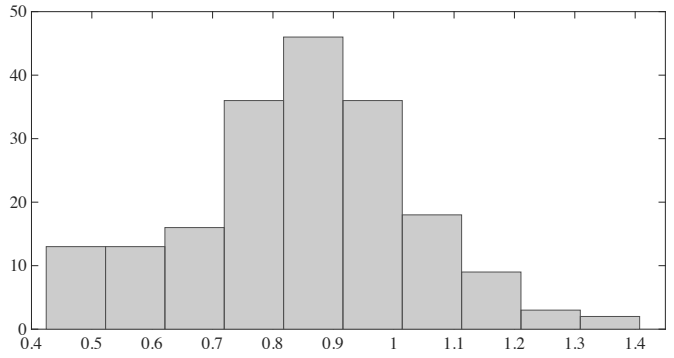
Figure 2 shows the average amplitude values histogram of the 600 targets extracted from the CARABAS II data set considering a 7×7 pixel window size. Figures 2(a), 2(b), and 2(c) present the histogram of the small, medium, and large target sizes, respectively. These histograms are representative of the targets' behavior of the three evaluated window sizes and, due to the limited amount of space of this paper, the other ones are omitted. The histogram of the small size targets given in Figures 2(a) shows a fragment and are distributed into two parts; hence, a unimodal distribution may not be the most accurate choice to fit this data. On the other hand, the targets with medium and large sizes can be modeled by traditional unimodal distributions.

Table I presents the AD test results for the evaluated distributions; the highlighted values are related to the AD statistic's smallest values. The value zero of AD test results means that the distribution is suitable for the data and the value one otherwise. The most accurate distribution for the tested data shows the smallest values of the AD test statistic.

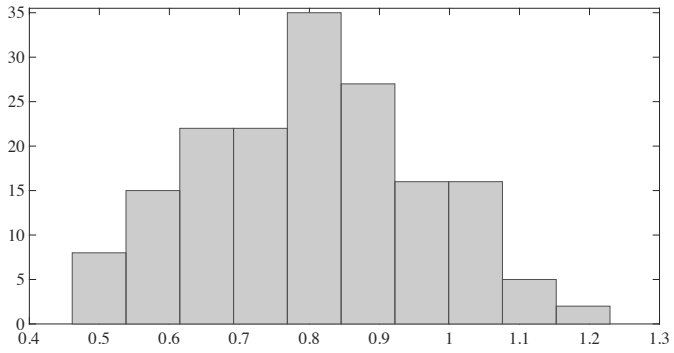
As a preliminary study, we manage to fit a Gaussian distribution in each small size target data part, since, as shown in Table I, this distribution is not suitable to fit the whole small size target data. However, we have that the Gaussian distribution can fit the data of each part, regardless of the evaluated window size; hence, we considered a Gaussian



(a) Small size



(b) Medium size



(c) Large size

Fig. 2. Target histograms considering the average amplitude value over a window of 7×7 pixels of each target.

mixture distribution to fit this data. As illustration, the data histogram of the average amplitude value over a window of 7×7 pixels of each target and the fitted density curves can be found in Figure 3(a). The Gaussian distribution may not be the most suitable distribution to model this data, since its support is the real numbers, and not just nonnegative values, such as the SAR image amplitude values. Thus, for future investigations, more accurate mixture models could be considered, considering suitable distributions for nonnegative values.

As the medium and large target sizes can be model by a unimodal distribution, we tested the Burr, Gamma, Log-normal, Rayleigh, Rician, and Weibull in each data set. As shown in Table I, considering windows of 5×5 and 7×7 pixels, the Burr, Gamma, Rician, and Weibull distribution can be suitable choices to model this data. For the window of 9×9 pixels, among the evaluated distributions, only the Rayleigh distribution is not suitable to fit this data. Furthermore, among the Burr, Gamma, Rician, and Weibull distributions and regardless of the window size, the smallest AD test statistic is related to the Burr distribution; for instance, the data histogram of the average amplitude value over a window of 7×7 pixels of each target and the fitted density curves can be found in Figure 3(b).

TABLE I

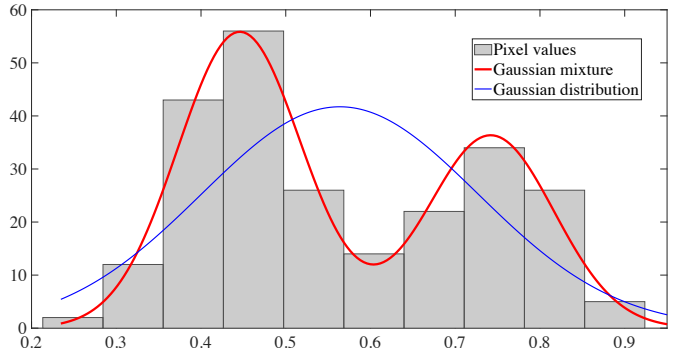
ANDERSON DARLING TEST RESULTS FOR THE BURR, GAMMA, GAUSSIAN, LOG-NORMAL, RAYLEIGH, RICIAN, AND WEIBULL DISTRIBUTIONS. THE HIGHLIGHTED VALUES ARE RELATED TO THE AD STATISTIC SMALLEST VALUES. AD TEST RESULT EQUAL ZERO MEANS THAT THE DISTRIBUTION IS SUITABLE FOR THE DATA AND ONE OTHERWISE

Test	Statistic	Result	Statistic	Result	Statistic	Result
Window size	5×5		7×7		9×9	
Small size						
Gaussian	8.0031	1	6.2261	1	5.1388	1
Gaussian (Part 1)	**0.3149**	0	**0.1954**	0	**0.3727**	0
Gaussian (Part 2)	**0.4795**	0	**0.3684**	0	**0.4278**	0
Medium size						
Burr	**0.9051**	0	**0.4709**	0	**0.4275**	0
Gamma	2.4254	0	1.7304	0	0.8226	0
Log-normal	3.4087	1	2.6432	1	1.3706	0
Rayleigh	28.8321	1	29.1844	1	29.2203	1
Rician	1.1305	0	0.6398	0	0.4579	0
Weibull	0.9242	0	0.5900	0	0.8730	0
Large size						
Burr	-	-	0.4882	0	0.3890	0
Gamma	**0.8569**	0	0.4825	0	0.5306	0
Log-normal	1.0096	0	0.7013	0	0.7879	0
Rayleigh	23.9868	1	28.3461	1	31.4313	1
Rician	0.9239	0	**0.4144**	0	**0.3220**	0
Weibull	1.0527	0	0.6807	0	0.5806	0

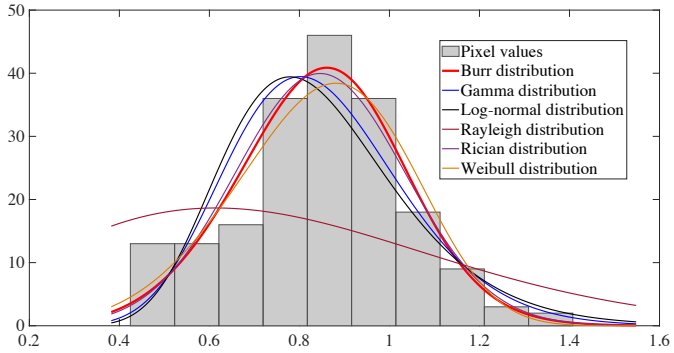
Based on the AD test results, among the tested distributions, the Rayleigh model is not suitable to fit the large size target data considering the three employed window sizes. Additionally, the data related to the window of 5×5 pixels is not fit by a Burr distribution with finite parameters. Based on the AD test statistic, the most accurate distribution to model the large size data is (i) the Gamma distribution, for the data obtained considering a window of 5×5 pixels; and (ii) the Rician model, for the windows of 7×7 and 9×9 pixels; for illustration, the data histogram of the average amplitude value over a window of 7×7 pixels of each target and the fitted density curves are shown in Figure 3(c). This particular experiment highlights the importance of considering an appropriate distribution to model each target and window size, aiming at obtaining a good data representation.

C. Computer-Simulated UWB VHF SAR Targets

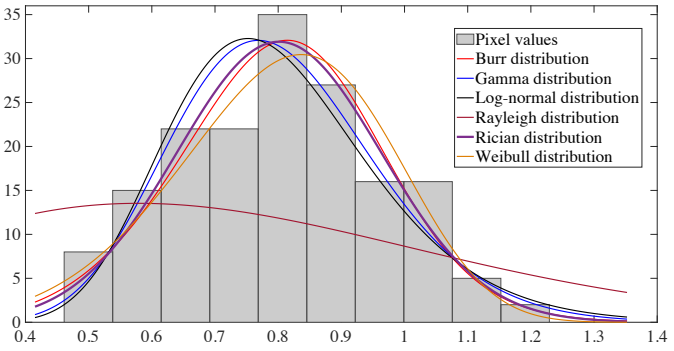
With the fitted distributions presented in Section III-B, we generate random values in order to simulate the modeled targets. Table II presents five examples of the average value of each real and computer-simulated target. For the three targets and window sizes, the real and computer-simulated targets



(a) Small size



(b) Medium size



(c) Large size

Fig. 3. Target histograms considering the average amplitude value over a window of 7×7 pixels of each target and the density curves of the fitted distributions.

show similar average values. Additionally, Table II shows the mean value of all 600 real targets for the evaluated window and target sizes; in general, the artificially generated target mean values are similar with the average of the 600 real ones, i.e., the simulated data keep the mean standard of the evaluated data set. In this preliminary study, we deal with the spatial dependence of pixels by ordering the values from the center to the edges of the window, i.e., the computer-simulated target central pixel has the highest magnitude, and so on.

Figure 4 shows an image example of real and computer-simulated targets considering the three evaluated target sizes and a window of 7×7 pixels. As discussed above, the computer-simulated targets kept statistical characteristics similar to the real ones. Finally, Figure 5 shows a CARABAS II zoomed image in the area with the real targets (25 white dots on the bottom) and the computer-simulated ones were added

TABLE II

REAL AND COMPUTER-SIMULATED TARGET MEAN VALUES

	Real	Simulated	Real	Simulated	Real	Simulated
Window size	5×5		7×7		9×9	
Small size						
Target 1	0.5882	0.5947	0.4237	0.4204	0.3535	0.3432
Target 2	0.5690	0.5355	0.4441	0.5279	0.4879	0.4563
Target 3	0.7430	0.7260	0.5079	0.5257	0.3535	0.3440
Target 4	0.5772	0.6151	0.5113	0.5389	0.3991	0.3686
Target 5	0.9771	0.9954	0.5650	0.5772	0.4653	0.4243
Medium size						
Target 1	1.6801	1.1702	0.8239	0.8348	0.6267	0.6139
Target 2	1.2622	1.1483	0.7934	0.8020	0.6832	0.6771
Target 3	1.5430	1.1523	0.8812	0.8641	0.7034	0.7149
Target 4	1.1310	1.1357	0.8449	0.8301	0.6267	0.6366
Target 5	1.1697	1.1909	0.8856	0.8891	0.6486	0.6591
Large size						
Target 1	1.0901	1.0797	0.8148	0.8161	0.6873	0.6548
Target 2	1.2425	1.1803	0.8463	0.8244	0.6816	0.6802
Target 3	1.0518	1.0515	0.7716	0.7887	0.6873	0.6744
Target 4	0.8395	0.9947	0.7967	0.7927	0.6359	0.6397
Target 5	1.3760	1.1368	0.8152	0.8182	0.6655	0.6631
Real target mean value (considering the 600 targets)						
Small size	0.7694		0.5638		0.4584	
Medium size	1.1581		0.8471		0.6600	
Large size	1.0626		0.8019		0.6496	

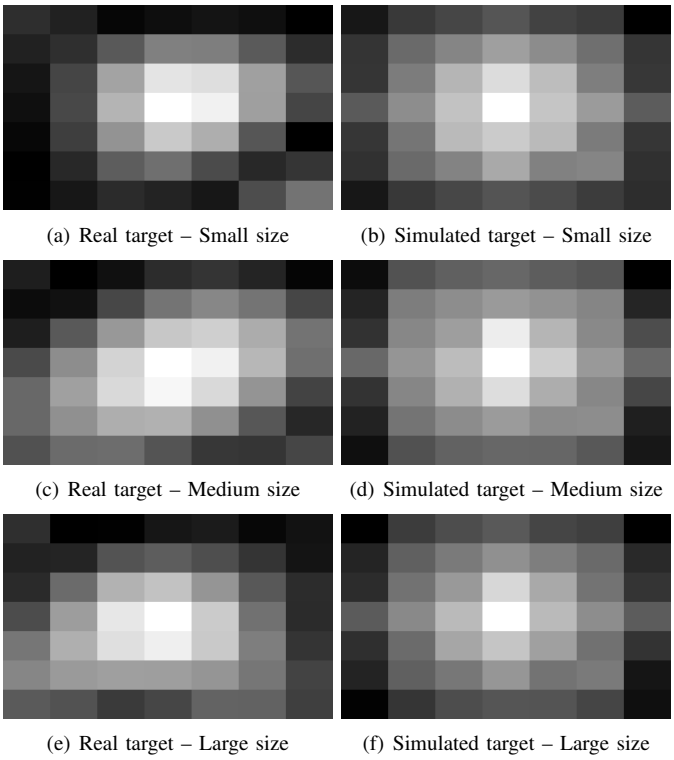
(15 white dots on the top). The difference in the computer-simulated and real target shapes are related to the real truck design (front and back part), while the computer-simulated targets, in these particular images, are a fixed window of 7×7 pixels. However, the statistical behavior of both targets is similar. Thus, in future studies, other morphological shapes can be considered to better represent the target morphologies.

Besides deriving the computer-simulated targets, the same statistical study can be applied in the images presented in [10], where a stack of images was used to obtain the ground scene true characteristics of the CARABAS II images, without targets. Hence, we can have a synthetic data set, considering computer-simulated ground scenes and targets.

IV. CONCLUSIONS

In this paper, we performed statistical models study of wavelength-resolution SAR image targets. For such, well-known statistical models, namely, Burr, Gamma, Gaussian, Log-normal, Rayleigh, Rician, and Weibull, and the AD test were considered. Based on this study, we proposed a preliminary method to obtain computer-simulated targets, which showed similar statistical behavior with the real ones.

For future studies, we aim at extending the proposed scheme, proposing a more suitable mixture distribution for the small size targets. Also, we intend to evaluate the flight heading and the target orientation in the pixel spatial distribution, and consequently, obtain different computer-simulated target morphological shapes. Additionally, the same process can be used to obtain computer-simulated ground scenes of SAR images. As a result, we expect to obtain trustworthy images with fully computer-simulated data.



(a) Real target – Small size (b) Simulated target – Small size

(c) Real target – Medium size (d) Simulated target – Medium size

(e) Real target – Large size (f) Simulated target – Large size

Fig. 4. Computer-simulated and real targets considering a window of 7×7 pixels.

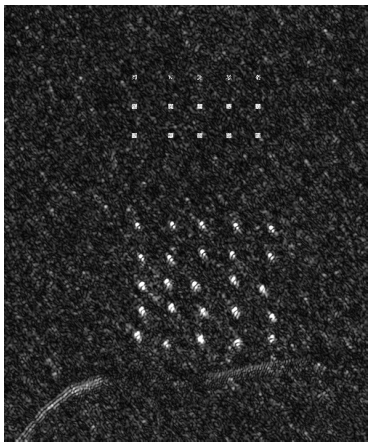


Fig. 5. CARABAS II SAR image with computer-simulated and real targets.

ACKNOWLEDGEMENTS

We gratefully acknowledge partial financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Edital Pró-Defesa IV, Brazil.

REFERENCES

[1] D. H. Hoekman and M. J. Quiriones, "Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 685–696, 2000.

[2] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1432–1445, 2007.

[3] H. Sportouche, J.-M. Nicolas, and F. Tupin, "Mimic capacity of fisher and generalized gamma distributions for high-resolution SAR image statistical modeling," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5695–5711, 2017.

[4] Y. Zheng, X. Zhang, B. Hou, and G. Liu, "Using combined difference image and k-means clustering for SAR image change detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 691–695, 2014.

[5] L. M. Ulander, M. Lundberg, W. Pierson, and A. Gustavsson, "Change detection for low-frequency SAR ground surveillance," *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 6, pp. 413–420, 2005.

[6] L. M. 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," in *Algorithms for Synthetic Aperture Radar Imagery XI*, vol. 5427. International Society for Optics and Photonics, 2004, pp. 259–271.

[7] L. Ulander, P.-O. Frolind, A. Gustavsson, H. Hellsten, T. Jonsson, B. Larsson, and G. Stenstrom, "Performance of the CARABAS-II VHF-band synthetic aperture radar," in *Geoscience and Remote Sensing Symposium, 2001. IGARSS'01. IEEE 2001 International*, vol. 1. IEEE, 2001, pp. 129–131.

[8] V. T. Vu, N. R. Gomes, M. I. Pettersson, P. Dammert, and H. Hellsten, "Bivariate gamma distribution for wavelength-resolution SAR change detection," *IEEE Transactions on Geoscience and Remote Sensing*, no. 99, pp. 1–9, 2018.

[9] M. Lundberg, L. M. Ulander, W. E. Pierson, and A. Gustavsson, "A challenge problem for detection of targets in foliage," in *Proceedings-Algorithms for Synthetic Aperture Radar Imagery XIII*, vol. 6237, 2006.

[10] B. G Palm, D. I Alves, M. I Pettersson, V. T Vu, R. Machado, R. J Cintra, F. M Bayer, P. Dammert, and H. Hellsten, "Wavelength-resolution SAR ground scene prediction based on image stack," *Sensors*, vol. 20, no. 7, p. 2008, 2020.

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[13] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[14] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1240–1244, 2019.

[15] U.S. Air Force, "The sensor data management system - SDMS," 2020, https://www.sdms.afrl.af.mil/index.php. [Online]. Available: https://www.sdms.afrl.af.mil/index.php

[16] D. Teguig, V. Le Nir, and B. Scheers, "Spectrum sensing method based on likelihood ratio goodness-of-fit test," *Electronics Letters*, vol. 51, no. 3, pp. 253–255, 2015.

[17] N. M. Razali, Y. B. Wah *et al.*, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[18] G. Wang, C. Lopez-Molina, G. V.-D. de Ulzurrun, and B. De Baets, "Noise-robust line detection using normalized and adaptive second-order anisotropic Gaussian kernels," *Signal Processing*, vol. 160, pp. 252–262, 2019.

[19] G. Wang, J. Zhu, and Z. Xu, "Asymptotically optimal one-bit quantizer design for weak-signal detection in generalized Gaussian noise and lossy binary communication channel," *Signal Processing*, vol. 154, pp. 207–216, 2019.

[20] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes," *The Annals of Mathematical Statistics*, pp. 193–212, 1952.

[21] R. B. D'Agostino, *Goodness-of-fit-techniques*. CRC press, 1986, vol. 68.

[22] Y. Wang, Y. Deng, W. Fei, R. Wang, H. Song, J. Wang, and N. Li, "Modified statistically homogeneous pixels' selection with multitemporal SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1930–1934, 2016.

[23] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in statistics*. Springer, 1992, pp. 66–70.