Low-Cost Robust Estimation for the Single-Look $\mathcal{G}_I^0$ Model Using the Pareto Distribution

Débora Chan, Andrea Rey, Juliana Gambini, and Alejandro C. Frery, *Senior Member, IEEE*

***Abstract*— The statistical properties of Synthetic Aperture Radar (SAR) image texture reveal useful target characteristics. It is well-known that these images are affected by speckle and prone to extreme values due to double bounce and corner reflectors. The $\mathcal{G}_I^0$ distribution is flexible enough to model different degrees of texture in speckled data. It is indexed by three parameters: α, related to the texture, γ, a scale parameter, and L, the number of looks. Quality estimation of α is essential due to its immediate interpretability. In this letter, we exploit the connection between the $\mathcal{G}_I^0$ and Pareto distributions. With this, we obtain six estimators that have not been previously used in the SAR literature. We compare their behavior with others in the noisiest case for monopolarized intensity data, namely single look case. We evaluate them using Monte Carlo methods for noncontaminated and contaminated data, considering convergence rate, bias, mean squared error, and computational time. We conclude that two of these estimators based on the Pareto law are the safest choices when dealing with actual data and small samples, as is the case of despeckling techniques and segmentation, to name just two applications. We verify the results with an actual SAR image.**

***Index Terms*— $\mathcal{G}_i^0$ distribution, speckle, parameter estimation.**

I. Introduction

SYNTHETIC APERTURE RADAR (SAR) is an active sensing instrument able to measure the roughness, electrical properties, and shape of the surface. It is widely used in environmental monitoring and evaluation of damages in natural catastrophes, among other applications. However, automatic SAR image interpretation is difficult due to the presence of speckle, making statistical modeling necessary.

Many statistical models have been proposed for monopolarized SAR image understanding. Gao [1] reviewed these models and considered the amplitude $\mathcal{G}^0$ distribution a breakthrough for its tractability and expressiveness.

Corresponding author: Juliana Gambini.

D. Chan and A. Rey are with Facultad Regional Buenos Aires Autónoma de Buenos, Universidad Tecnológica Ciudad, Aires C1179AAP, Argentina (e-mail: mchan@frba.utn.edu.ar; arey@frba.utn.edu.ar).

J. Gambini is with Departamento de Ingeniería Informática, Instituto Tecnológico de Buenos Aires, Buenos Aires C1437FBG, Argentina, and also with the Depto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero, Pcia. de Buenos, Aires 1674, Argentina (e-mail: mgambini@itba.edu.ar).

A. C. Frery is with the Laboratório de Computacao Científica e Análise Numérica, Universidade Federal de Alagoas, Maceió 57072-900, Brazil, and also with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an 710071, China (e-mail: acfrery@laccan.ufal.br).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Many applications employ the intensity version, the $\mathcal{G}_I^0$ law, because it does not involve special functions other than the Euler's Gamma, and the ability to describe a wide variety of targets [2]. Three parameters index the $\mathcal{G}_I^0$ distribution: α, related to the texture, γ, a scale parameter, and L, the number of looks which is related to the signal-to-noise ratio. The last parameter may be known or estimated for the whole image, while the others describe local characteristics.

The texture is a pivotal parameter in SAR image interpretation; thus, precision and accuracy in the estimation of α are basilar. This inference becomes critical when only a few samples are available, and when they are prone to contamination, as with local statistical filters and segmentation.

We obtain estimators that have not been used by the SAR community through a connection between the $\mathcal{G}_I^0$ and Pareto laws. We analyze these estimation methods in terms of bias, mean square error, convergence, and computational time. We consider two situations: pure data and contaminated observations. The latter describes the situation of a few large values, e.g., from a double bounce or a corner reflector, with respect to the background mean.

We evaluate maximum likelihood (ML), penalized ML (PML), likelihood moments (LMs), probability weighted moments (PWMs), maximum goodness of fit (MGF) with different statistics, and minimum density power divergence (MPD) estimators in the single look case ($L = 1$).

The ML estimator is widely used because of its asymptotic properties, even though it has bias, and robustness problems with small samples. Several attempts have been made to reduce ML bias using analytic [3] and bootstrap methods [4], [5]. Frery *et al.* [6] proposed a technique to correct its tendency to diverge with small samples.

Robustness is another desired property. Among the possible deviations from the ideal situation of independent identically distributed (i.i.d.) deviates, extreme values are frequent in SAR imagery due to, for instance, corner reflectors and other sources of double bounce. Such departures from the basic model may compromise the performance of, for instance, despeckling filters and segmentation techniques. Among the robust proposals, M- and AM-estimators proved to be reliable in the presence of corner reflectors [7], [8].

Gambini *et al.* [9] formulated a nonparametric method which consists in minimizing the triangular distance between the $\mathcal{G}_I^0$ density probability function and an empirical distribution of the data computed with asymmetric kernels. This proposal is robust but has high computational cost, and it fails in the single look case. Wang *et al.* [10] developed a robust estimator based on the random weighting method.

They evaluated its performance for different values of L under contamination, and they concluded that the bigger the number of looks, the smaller the percentage of no convergence. For these reasons, in this work, we study estimation methods for $L = 1$, the noisiest situation.

The $\mathcal{G}_I^0$ law is a Generalized Pareto type II distribution [11] when $L = 1$. This law has been verified and studied in several fields because of its flexibility to model different phenomena. We take advantage of this fact by using estimators whose good properties (bias corrected [3], low computational cost and asymptotic efficiency [12]) have already been assessed but that have not been used by the SAR community.

Our study compares parameter estimation techniques according to their computational cost, convergence rate, bias, and mean squared error (MSE) for data without contamination using Monte Carlo methods. We use the influence function to quantify the estimators' robustness. We then apply these methods to actual data with a corner reflector.

This letter unfolds as follows. In Section II some $\mathcal{G}_I^0$ distribution properties are recalled, and the state of art of the selected estimators is revised. Sections III-A and III-B discuss the results obtained with simulation. Section III-C shows applications to actual data. Section IV discusses remarks and presents recommendations for practical applications. Finally, the Supplementary Material provides definitions, code, and other details.

II. State of Art

A. Single Look $\mathcal{G}_I^0$ Model for Intensity Data

The $\mathcal{G}_I^0$ distribution is characterized by the following probability density function:

$$f_{\mathcal{G}_I^0}(z) = \frac{L^L \Gamma(L-\alpha)}{\gamma^{\alpha}\Gamma(-\alpha)\Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}} \mathbb{1}_{\mathbb{R}_+}(z) \tag{1}$$

where $-\alpha, \gamma > 0$ and $L \geq 1$. We are interested in the noisiest case, which corresponds to $L = 1$, called the single look case. The probability density function becomes

$$f_{\mathcal{G}_I^0}(z) = \frac{-\alpha}{\gamma}\left(\frac{z}{\gamma} + 1\right)^{\alpha-1} \mathbb{1}_{\mathbb{R}_+}(z). \tag{2}$$

The r-order moments for $L = 1$ are

$$\mathrm{E}(Z^r) = \gamma^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)}\Gamma(1 + r) \tag{3}$$

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

The generalized Pareto type II distribution, $GP_{\mathrm{II}}(\mu, \sigma, \beta)$ has the following probability density function:

$$f_{GP_{\mathrm{II}}}(z) = \frac{\beta}{\sigma}\left(1 + \frac{z-\mu}{\sigma}\right)^{-\beta-1} \mathbb{1}_{[\mu,+\infty)}(z) \tag{4}$$

so the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution is a particular case of this distribution for $\mu = 0$, $\sigma = \gamma$ and $\beta = -\alpha$. Using this fact, we take advantage of the extensive existing bibliography on parameter estimation under the Pareto distribution, applying the best estimators to solve our problem.

B. Parameter Estimation Methods

In this section, we comment some of the characteristics of the methods reviewed in this work. Their implementation and other details can be seen in the Supplementary Material.

TABLE I
Abbreviations List

Abbreviation	Estimator	Notation
ML	Maximum Likelihood	$\widehat{\alpha}_{\mathrm{ML}}$
PML	Penalized Maximum Likelihood	$\widehat{\alpha}_{\mathrm{PML}}$
Mom	Moments	$\widehat{\alpha}_{\mathrm{Mom}}$
PWM	Penalized Weighted Moments	$\widehat{\alpha}_{\mathrm{PWM}}$
LM	Likelihood Moments	$\widehat{\alpha}_{\mathrm{LM}}$
Med	Median	$\widehat{\alpha}_{\mathrm{Med}}$
MPD	Minimum Power Density Divergence	$\widehat{\alpha}_{\mathrm{MPD}}$
MGF	Maximumm Goodness of Fit	$\widehat{\alpha}_{\mathrm{MGF}}$
ADR	MGF with Anderson Darling Right Tail	$\widehat{\alpha}_{\mathrm{ADR}}$

Table I shows the methods reviewed in this work, along with their abbreviations.

Grimshaw [13] studied the properties of the ML estimator under the generalized Pareto distribution (GPD) model. It is asymptotically consistent, efficient, and normal, but in many cases, it has not an explicit solution, and it diverges for small samples.

Coles and Dixon [14] proposed maximizing the log-likelihood plus a penalization function, the PML method, in order to reach convergence for small samples. For large sample sizes, PML inherits optimal properties from ML, avoiding its limitations in small ones.

Method of Moments (Mom) estimators have asymptotic normality, asymptotic efficient normality and, in some cases, explicit solution. They can be generalized to PWM estimators, as detailed in the Supplementary Material. Hosking and Wallis [15] discussed some properties of Mom and PWM estimators for the GPD distribution. They compared the performance of Mom, ML, and PWM and observed that Mom is asymptotically normal, but it frequently does not converge, and it is sensitive to outliers. They showed that PWM is an alternative to ordinary moments with advantages for small sample sizes but with low asymptotic efficiency.

Zhang [12] proposed the LMs estimator by combining ML and Mom techniques. The solution always exists and is efficient and asymptotically normal.

Peng and Welsh [16] proposed the median (Med) estimator, as a robust alternative, by solving an equation that relates the sample and population likelihood score medians. It is robust because of its bounded influence function, and it is asymptotically normal, but in many cases, it does not have good performance [17] being, thus, only advisable under the presence of outliers.

The maximum power density divergence estimator is another robust alternative. It has bounded influence function and is indexed by a positive constant ω, which controls the tradeoff between efficiency and robustness. As ω rises, robustness increases and efficiency decreases, becoming ML when $\omega = 0$. Juárez and Schucany [17] proved that ML is more efficient for noncontaminated data but MPD outperforms it under contamination with grossly outlying observations.

Luceño [18] proposed MGF estimators. He showed their consistency and asymptotic efficiency. The author proved its superiority for heavy-tailed data.

Fig. 1 summarizes the main characteristics of the considered parameter estimation methods, where "Asymp. Norm,"

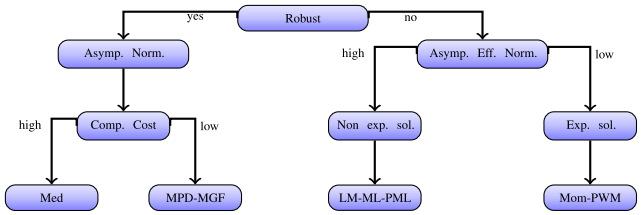


Fig. 1. Parameter estimation method characteristics.

"Asymp. Eff. Norm." and "Exp. Sol." mean asymptotic normality, asymptotic efficient normality, and explicit solution, respectively.

The GPD is the limiting law of normalized excesses over a threshold [19]. Thus, the choice of threshold is crucial for accurate estimation. We assessed the performance of the estimators above with the following threshold selection methods: 1) u_0 which considers the whole sample; 2) u_{q10} which uses the 90% largest values; 3) u_{q20} which considers the 80% largest values; 4) u_{Hill} based on the Hill plot; 5) u_{AD} which is an automated threshold selection based on the p-values of the attention display (AD) goodness of fit test.

In order to decide the most suitable threshold for each estimator, we generated 1000 samples of sizes 25, 49, and 81, for all combinations of $\alpha \in \{-8, -5, -2\}$ and $\gamma \in \{0.1, 1, 10\}$. We concluded that u_{q10} is the best threshold for $n = 49$, for MPD and ML methods. In other cases, the choice is u_0. Also, we compared the following goodness-of-fit test statistics: Anderson Darling left tail, Kolmogorov Smirnov, Cramer von Mises, square Anderson Darling, and Anderson Darling right tail. We obtained the best results with the latter, so this is the only method we present here. We do not report the results with Mom and Med estimators due to their relatively poor performance.

III. Simulation Experiments and Results

We compared the following estimation methods: MGF with Anderson Darling right tail (ADR), LM, ML, MPD, PML, and PWM by their MSE, convergence rate, bias and computational time, for noncontaminated and contaminated data. We also present the results of applying the methods to actual data.

The Supplementary Material shows the numerical evidence that led to the conclusions we present in Sections III-A–III-C.

A. Noncontaminated Data

We considered 1000 samples from the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution, of sizes $\{25, 49, 81, 121, 500\}$ without contamination combining the parameter values $\alpha \in \{-8, -5, -2\}$ and $\gamma \in \{0.1, 1, 10, 100\}$. We obtained the samples following the guidelines presented in [20].

For small samples, we observed the best convergence rate for ADR. Except for PML, the rest of the considered estimators reach good convergence rate with large samples and an adequate one with small samples. LM outperforms the other methods for lower texture values. See the convergence rate for high, medium, and low texture and for $\gamma = \{1, 10\}$ in Fig. SM-1 (Supplementary Material).

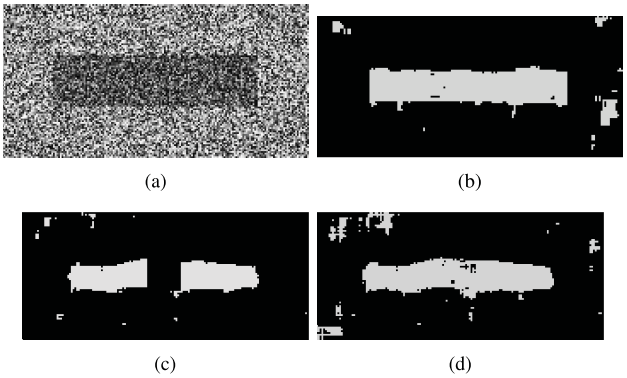


Fig. 2. Support vector machine (SVM) image classification with and without corner reflector. (a) Simulated image with two regions. (b) SVM image classification using PWM Method, accuracy: 0.82. (c) SVM image classification using PWM with a corner reflector, accuracy: 0.77. (d) SVM image classification using MPD with a corner reflector, accuracy: 0.81.

For low texture values, all the methods underestimate and for high texture values, they overestimate. We note that, for $\alpha = -8$, PML outperforms the others even for small samples. PWM enhances its performance as the texture value decreases. LM stands out in some cases; meanwhile, all estimators achieve good accuracy in sample sizes larger than 121. Figs. SM-2 and SM-3 (Supplementary Material) show the bias according to the texture, and that the performance is the same for all the candidates. The lack of monotonicity of some curves is due to the randomness and to the heavytailedness of the model.

Multiple comparisons of Tukey honestly significant difference (HSD) test point out that the PWM method is the fastest. Fig. SM-4 (Supplementary Material) shows the time consumed in milliseconds by each method for 1000 samples of size 500 for all parameter combinations.

B. Contaminated Data

The presence of outliers may cause gross errors in the parameter estimation. We generated contaminated random samples in order to assess the estimators under contamination.

Let B be a Bernoulli random variable with a probability of success $0 < \epsilon \ll 1$. We are interested in a contaminated random variable $Z = BC + (1 - B)W$, where $C \in \mathbb{R}_+$ and $W \sim \mathcal{G}_I^0(\alpha, \gamma, 1)$.

As a way of measuring the impact of this contamination, we constructed stylized empirical influence functions (SEIFs) [8] for samples of sizes $n \in \{25, 49, 81, 121\}$, for each estimator considering all parameter combinations. Such function measures how much a single observation is able to drag the estimate away from its ideal value. Fig. SM-5 (Supplementary Material) shows such functions for $\alpha = -5$ and $\gamma = 100$. With this, the expected value of the background is 25; the contamination C spans from 25 to 1000. We fixed $\epsilon = (n + 1)^{-1}$, where n is the sample size.

With this approach, we verified that: 1) MPD is the least sensitive estimator to large contamination and 2) the most sensitive is PWM. Juárez and Schucany [17] proved that the influence of outliers over MPD is bounded.

We carried out the following experiment in order to illustrate the importance of using more than one estimator.

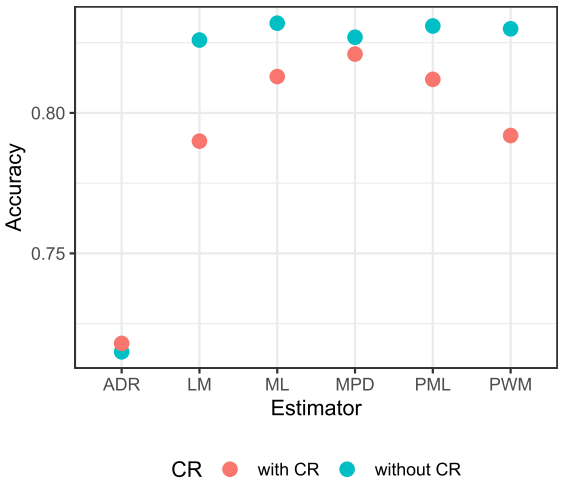
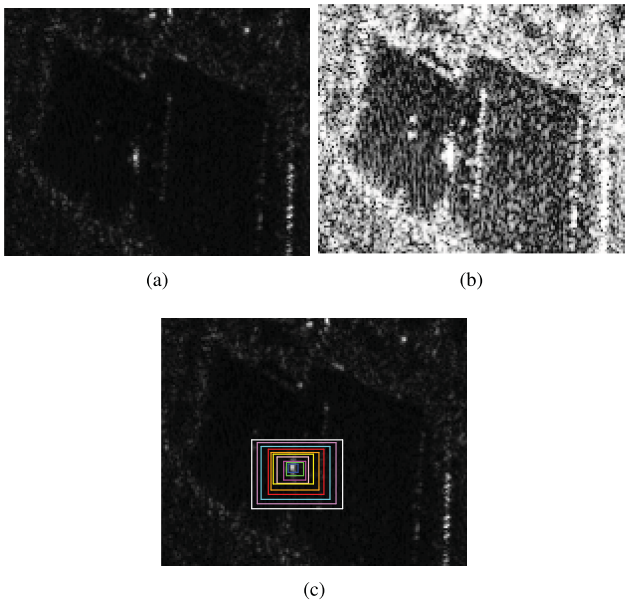


Fig. 3. Accuracy classification values with and without corner reflector.



(a) (b) (c)

Fig. 4. (a) Single-look E-SAR image with a corner reflector, used to estimate the α-parameter, (b) its equalized version, and (c) ten Regions of Interest of different sizes used to estimate the texture parameter.

1) A simulated SAR image is generated sampling from the $\mathcal{G}_I^0(-5,1,1)$ and $\mathcal{G}_I^0(-2,1,1)$ distributions for the central region and for the edge zone, respectively; one of these images is shown in Fig. 2(a). The mean values are 4 and 1.
2) An image of estimates is built using the PWM method, then a classification is performed using the SVM supervised method and a linear kernel. Fig. 2(b) shows the result.
3) We add an outlier at the center of the original image to obtain a contaminated image. The outlier is one pixel with value equal to 100.
4) We repeat the procedure: estimate by PWM, then SVM classification. Fig. 2(c) shows the result.
5) We make an image of estimates with MPD, then classify it with the same SVM method and obtain the results shown in Fig. 2(d).

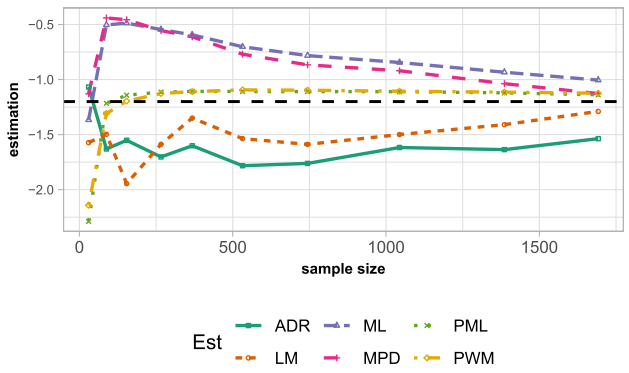
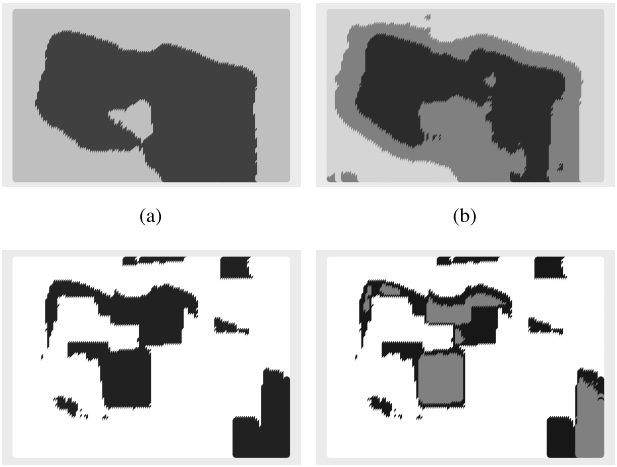


Fig. 5. Texture estimates using the samples from Fig. 4(c).



(a) (b) (c) (d)

Fig. 6. Two classes segmentation using α parameter estimation and k-means classification method. (a) k-means classification in two regions using LM method. (b) k-means classification in three regions using LM method. (c) k-means classification in two regions using ML method. (d) k-means classification in three regions using ML method.

It can be observed that the parameter estimation method depends strongly on the structure of the image. In the above experiment, we notice that in the absence of the corner reflector, the PWM method provides good classification results, but in the presence of the corner reflector it is necessary to use another method. The classification accuracy is shown in Fig. 3.

Fig. 3 shows the accuracy of the classification with each estimation method, with and without corner reflector. MPD has one of the basilar properties of a good robust estimator: being acceptable when there is no contamination, and not producing unacceptable results under the presence of outlying data.

C. Actual SAR Data

Fig. 4(a) shows an intensity single-look L-band HH polarization E-SAR image with a corner reflector used to compute the estimates. Fig. 4(b) shows the equalized SAR image from Fig. 4(a). It can be observed the complex texture structure of the image. Fig. 4(c) shows the regions used for estimating the texture parameter. The estimates are presented in Fig. 5 where

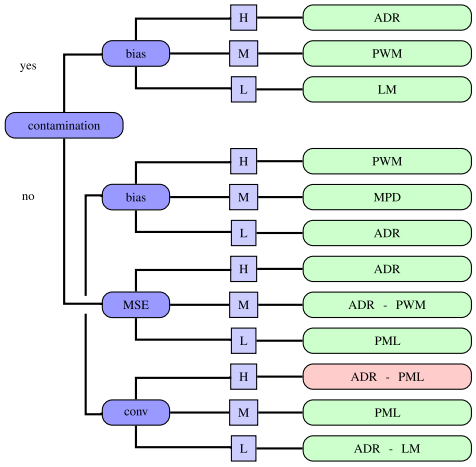


Fig. 7. Estimation performance by texture: high (H), medium (M) and low (L). Green and red blocks mean appropriate and poor performance, respectively.

the dashed black line shows the considered true value as the asymptotic limit of the ML estimator.

All estimates approach the hypothesized true value, but at very different rates; some underestimate it (ADR and LM), while others overestimate it (ML, MPD, PML, PWM). Besides the rate of convergence, the behavior is quite different from that of small samples. The following application shows the impact of such behavior in a classification process.

Fig. 6 shows the results of applying a two-step segmentation to the image in Fig. 4(a): first a parameter estimation with LM and ML methods using windows of size 21×21 pixels, followed by a k-means classification with $k = 2, 3$. The classification obtained with LM estimates outlines quite well the structure observed in Fig. 4(a). This is due to the robustness of the estimation procedure before the presence of a corner reflector.

IV. Conclusion

In this letter, we compared six parameter estimation methods with and without contamination. We evaluate MSE, convergence rate, bias, and computational time. The i.i.d. model assumed for noncontaminated data does not hold in case of the presence of a corner reflector, which can be considered a contamination. We analyzed the impact of such contamination by means of stylized empirical influence curves. Hence, when the image is susceptible to having this type of contamination, it is recommended to consider robust techniques.

ML and MPD converge for large samples but present a significant bias in small samples, so we consider that PWM, PML and LM outperform them in this case. Also, PWM consumes the lowest computational time.

In addition, we observed that the performance of the estimation methods strongly depends on the texture characteristics of the region of interest. For this reason, it is very important to use more than one estimator, so as to improve the results of the automatic interpretation of actual images. Fig. 7 shows a scheme summarizing our recommendations.

Appendix

Simulations were performed using the R language and environment for statistical computing version 3.3 [21], in a computer with processor Intel Core, i7-4790K CPU 4-GHz, 16-GB RAM, system type 64-bit operating system.

References

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

[2] M. E. Mejail, J. C. Jacobo-Berlles, A. C. Frery, and O. H. Bustos, "Classification of SAR images using a general and tractable multiplicative model," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3565–3582, 2003.

[3] D. E. Giles, H. Feng, and R. T. Godwin, "Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution," *Commun. Statist. Theory Methods*, vol. 45, no. 8, pp. 2465–2483, 2016.

[4] K. L. P. Vasconcellos, A. C. Frery, and L. B. Silva, "Improving estimation in speckled imagery," *Comput. Statist.*, vol. 20, no. 3, pp. 503–519, 2005.

[5] M. F. da Silva, F. Cribari-Neto, and A. C. Frery, "Improved likelihood inference for the roughness parameter of the $\mathcal{G}_A^0$ distribution," *Environmetrics*, vol. 19, no. 4, pp. 347–368, 2008.

[6] A. C. Frery, F. Cribari-Neto, and M. O. de Souza, "Analysis of minute features in speckled imagery with maximum likelihood estimation," *EURASIP J. Adv. Signal Process.*, vol. 16, Dec. 2004, Art. no. 375370.

[7] O. H. Bustos, M. M. Lucini, and A. C. Frery, "M-estimators of roughness and scale for $\mathcal{G}_A^0$-modelled SAR imagery," *EURASIP J. Adv. Signal Process.*, vol. 1, no. 1, Dec. 2002.

[8] H. Allende, A. C. Frery, J. Galbiati, and L. Pizarro, "M-estimators with asymmetric influence functions: The G_A^0 distribution case," *J. Stat. Comput. Simul.*, vol. 76, no. 11, pp. 941–956, 2006.

[9] J. Gambini, J. Cassetti, M. M. Lucini, and A. C. Frery, "Parameter estimation in SAR imagery using stochastic distances and asymmetric kernels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 365–375, Jan. 2015.

[10] C.-H. Wang, X.-B. Wen, and H.-X. Xu, "A robust estimator of parameters for G_A^0-modeled SAR imagery based on random weighting method," *EURASIP J. Adv. Signal Process.*, vol. 22, p. 22, Feb. 2017.

[11] D. Chan, J. Cassetti, and A. C. Frery, "Texture parameter estimation in monopolarized SAR imagery, for the single look case, using extreme value theory," in *Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3266–3269.

[12] J. Zhang, "Likelihood moment estimation for the generalized Pareto distribution," *Austral. New Zealand J. Statist.*, vol. 49, no. 1, pp. 69–77, 2007.

[13] S. D. Grimshaw, "Computing maximum likelihood estimates for the generalized Pareto distribution," *Technometrics*, vol. 35, no. 2, pp. 185–191, 1993. doi: 10.2307/1269663.

[14] S. G. Coles and M. J. Dixon, "Likelihood-based inference for extreme value models," *Extremes*, vol. 2, no. 1, pp. 5–23, 1999.

[15] J. R. M. Hosking and J. R. Wallis, "Parameter and quantile estimation for the generalized Pareto distribution," *Technometrics*, vol. 29, no. 3, pp. 339–349, 1987.

[16] L. Peng and A. Welsh, "Robust estimation of the generalized Pareto distribution," *Extremes*, vol. 4, no. 1, pp. 53–65, 2001.

[17] S. F. Juárez and W. R. Schucany, "Robust and efficient estimation for the generalized Pareto distribution," *Extremes*, vol. 7, no. 3, pp. 237–251, 2004.

[18] A. Luceño, "Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators," *Comput. Statist. Data Anal.*, vol. 51, pp. 904–917, 2006.

[19] J. Pickands, III, "Statistical inference using extreme order statistics," *Ann. Statist.*, vol. 3, no. 1, pp. 119–131, 1975.

[20] D. Chan, A. Rey, J. Gambini, and A. C. Frery, "Sampling from the G_I^0 distribution," *Monte Carlo Methods Appl.*, vol. 24, no. 4, pp. 271–287, 2018.

[21] R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. [Online]. Available: https://www.R-project.org/