Parameter Estimation in SAR Imagery Using Stochastic Distances and Asymmetric Kernels

Juliana Gambini, Julia Cassetti, María Magdalena Lucini, and Alejandro C. Frery, *Senior Member, IEEE*

Abstract—In this paper, we analyze several strategies for the estimation of the roughness parameter of the $\mathcal{G}_I^0$ distribution. It has been shown that this distribution is able to characterize a large number of targets in monopolarized synthetic aperture radar (SAR) imagery, deserving the denomination of "Universal Model." It is indexed by three parameters: 1) the number of looks (which can be estimated in the whole image); 2) a scale parameter; and 3) the roughness or texture parameter. The latter is closely related to the number of elementary backscatters in each pixel, one of the reasons for receiving attention in the literature. Although there are efforts in providing improved and robust estimates for such quantity, its dependable estimation still poses numerical problems in practice. We discuss estimators based on the minimization of stochastic distances between empirical and theoretical densities and argue in favor of using an estimator based on the triangular distance and asymmetric kernels built with inverse Gaussian densities. We also provide new results regarding the heavy-tailedness of this distribution.

Index Terms—Feature extraction, image texture analysis, speckle, statistics, synthetic aperture radar (SAR).

I. INTRODUCTION

THE STATISTICAL modeling of the data is essential in order to interpret synthetic aperture radar (SAR) images. Speckled data have been described under the multiplicative model using the $\mathcal{G}$ family of distributions, which is able to describe rough and extremely rough areas better than the $\mathcal{K}$ distribution [1], [2]. The survey article [3] discusses in detail several statistical models for this kind of data.

Under the $\mathcal{G}$ model, different degrees of roughness are associated with different parameter values; therefore, it is of paramount importance to have high quality estimators. Several works have been devoted to the subject of improving estimation with two main venues of research, namely, analytic and resampling procedures.

J. Gambini is with the Instituto Tecnológico de Buenos Aires, C1106 ACD Buenos Aires, Argentina, and also with the Depto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero, Pcia. de Buenos Aires, Argentina (e-mail: juliana.gambini@gmail.com).

J. Cassetti is with the Instituto de Desarrollo Humano, Universidad Nacional de Gral. Sarmiento, 1613 Pcia. de Buenos Aires, Argentina (e-mail: julia.cassetti@gmail.com).

M. M. Lucini is with the Facultad de Ciencias Exactas, Naturales y Agrimensura, Universidad Nacional del Nordeste, 3400 Corrientes, Argentina (e-mail: magdalenalucini@gmail.com).

A. C. Frery is with the LaCCAN, Universidade Federal de Alagoas, 57072-900 Maceió, Brazil (e-mail: acfrery@gmail.com).

The analytic approach was taken by Vasconcellos *et al.* [4] who quantified the bias in the estimation of the roughness parameter of the $\mathcal{G}_A^0$ distribution by maximum likelihood (ML). They proposed an analytic change for improved performance with respect to bias and mean-squared error. Also, Silva *et al.* [5] computed analytic improvements for that estimator. Such approaches reduce both the bias and the mean-squared error of the estimation, at the expense of computing somewhat cumbersome correction terms, and yielding estimators whose robustness is largely unknown.

Cribari-Neto *et al.* [6] compared several numerical improvements for that estimator using bootstrap. Again, the improvement comes at the expense of intensive computing and with no known properties under contamination. Allende *et al.* [7] and Bustos *et al.* [8] also sought for improved estimation, but seeking the robustness of the procedure. The new estimators are resistant to contamination, and in some cases, they also improve the mean-squared error, but they require dealing with influence functions and asymptotic properties not always immediately available to remote sensing practitioners.

A common issue in all the aforementioned estimation procedures, including ML, and to those based on fractional moments [1] and log-cumulants [9]–[11] is the need of iterative algorithms for which there is no granted convergence to global solutions. Such lack of convergence usually arises with small samples, precluding the use of such techniques in, e.g., statistical filters. Frery *et al.* [12] and Pianto and Cribari-Neto [13] proposed techniques that aim at alleviating such issue, at the cost of additional computational load.

The main idea of this work is to develop an estimation method for the $\mathcal{G}_I^0$ model with good properties (as measured by its bias, the mean-squared error and its ability to resist contamination) even with samples of small and moderate size and low computational cost. In order to achieve this task, we propose minimizing a stochastic distance between the fixed empirical evidence and the estimated model.

Shannon proposed a divergence between two density functions as a measure of the relative information between the two distributions. Divergences were studied by Kullback and Leibler and by Rényi [14], among others. These divergences have multiple applications in signal and image processing [15], medical image analysis diagnosis [16], and automatic region detection in SAR imagery [17]–[20]. Liese and Vajda [21] provide a detailed theoretical analysis of divergence measures.

Cassetti *et al.* [22] compared estimators based on the Hellinger, Bhattacharyya, Rényi, and Triangular distance with the ML estimator. They presented evidence that the Triangular

distance is the best choice for this application, but noticed that histograms led to many numerical instabilities. This work presents improvements with respect to those results in the following regards: we assess the impact of contamination in the estimation, we employ kernels rather than histograms and we compare estimators based on the Triangular distance with the ML, fractional moments, and log-cumulants estimators. Among the possibilities for such estimate, we opted for employing asymmetric kernels.

Kernels have been extensively applied with success to image processing problems as, e.g., object tracking [23]. Kernels with positive support, which are of particular interest for our work, were employed in [24] and [25]. The general problem of using asymmetric kernels was studied in [26].

In [27], the authors demonstrate, through extensive modeling of real data, that the SAR images are best described by families of heavy-tailed distributions. We provide new results regarding the heavy-tailedness of the $\mathcal{G}_I^0$ distribution; in particular, we show that it has heavy tails with tail index $1 - \alpha$.

This paper unfolds as follows. Section II recalls the main properties of the $\mathcal{G}_I^0$ modeol and the ML, the $\frac{1}{2}$-moment and the log-cumulants methods for parameter estimation. Estimation with stochastic distances is presented in Section III, including different ways of calculating the estimate of the underlying density function and the contamination models employed to assess the robustness of the procedure. Section IV presents the main results. Section V discusses the conclusion.

II. $\mathcal{G}_I^0$ MODEL

The return in monopolarized SAR images can be modeled as the product of two independent random variables, corresponding to 1) the backscatter X; and 2) the speckle noise Y. Similarly, $Z = XY$ represents the return in each pixel under the multiplicative model.

For monopolarized data, speckle is modeled as a Γ distributed random variable, with unitary mean and shape parameter $L \geq 1$, the number of looks, while the backscatter is considered to obey a reciprocal of Gamma law. This gives rise to the $\mathcal{G}_I^0$ distribution for the return. Given the mathematical tractability and descriptive power of the $\mathcal{G}_I^0$ distribution for intensity data [2], [28], it represents an attractive choice for SAR data modeling.

The density function for intensity data is given by

$$f_{\mathcal{G}_I^0}(z) = \frac{L^L \Gamma(L - \alpha)}{\gamma^\alpha \Gamma(-\alpha)\Gamma(L)} \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}} \quad (1)$$

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

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

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

With the double purpose of simplifying the calculations and making the results comparable, in the following, we choose the scale parameter such that $E(Z) = 1$, which is given by $\gamma^* = -\alpha - 1$.

One of the most important features of the $\mathcal{G}_I^0$ distribution is the interpretation of the α parameter, which is related to the roughness of the target. Values close to zero (typically above -3) suggest extreme textured targets, as urban zones. As the value decreases, it indicates regions with moderate texture (usually $\alpha \in [-6, -3]$), as forest zones. Textureless targets, e.g., pasture, usually produce $\alpha \in (-\infty, -6)$. This is the reason why the accuracy in the estimation of α is so important.

Let $\boldsymbol{z} = (z_1, \ldots, z_n)$ be a random sample of n independent draws from the $\mathcal{G}_I^0$ model.

Assuming $\gamma^* = -\alpha - 1$, the ML estimator for the parameter α, namely $\widehat{\alpha}_{\mathrm{ML}}$, is the solution of the following nonlinear equation:

$$\Psi^0(\widehat{\alpha}_{\mathrm{ML}}) - \Psi^0(L - \widehat{\alpha}_{\mathrm{ML}}) - \log(1 - \widehat{\alpha}_{\mathrm{ML}})$$
$$+ \frac{\widehat{\alpha}_{\mathrm{ML}}}{1 - \widehat{\alpha}_{\mathrm{ML}}} + \frac{1}{n} \sum_{i=1}^n \log(1 - \widehat{\alpha}_{\mathrm{ML}} + Lz_i)$$
$$- \frac{\widehat{\alpha}_{\mathrm{ML}} - L}{n} \sum_{i=1}^n \frac{1}{1 - \widehat{\alpha}_{\mathrm{ML}} + Lz_i} = 0$$

where $\Psi^0(\cdot)$ is the digamma function. Some of the issues posed by the solution of this equation have been discussed, and partly solved, in [12].

Fractional moments estimators have been widely used with success [1], [29]. Using $r = 1/2$ in (2), one has to solve

$$\frac{1}{n} \sum_{i=1}^n \sqrt{z_i} - \sqrt{\frac{-\widehat{\alpha}_{\mathrm{Mom12}} - 1}{L}} \frac{\Gamma(-\widehat{\alpha}_{\mathrm{Mom12}} - \frac{1}{2})}{\Gamma(-\widehat{\alpha}_{\mathrm{Mom12}})} \frac{\Gamma(L + \frac{1}{2})}{\Gamma(L)} = 0. \quad (3)$$

Estimation based on log-cumulants is gaining space in the literature due to its nice properties and good performance [9]–[11]. Following Tison *et al.* [30], the main second kind statistics can be defined as follows.

1) First second kind characteristic function: $\phi_x(s) = \int_0^\infty u^{s-1} p_x(u) du$.
2) Second second kind characteristic function: $\psi_x(s) = \log \phi_x(s)$.
3) First-order second kind characteristic moment

$$\tilde{m}_r = \frac{d\phi_x(s)}{ds}\bigg|_{s=1}. \quad (4)$$

4) First-order second kind characteristic cumulant (log-cumulant)

$$\tilde{k}_r = \frac{d\psi_x(s)}{ds}\bigg|_{s=1}. \quad (5)$$

If $p_x(u) = f_{\mathcal{G}_I^0}(u)$, then we have

1) $\phi_x(s) = \frac{\left(\frac{L}{\gamma}\right)^{1-s} \Gamma(-1+L+s)\Gamma(1-s-\alpha)}{\Gamma(L)\Gamma(-\alpha)}$
2) $\widetilde{m}_1 = \frac{d\phi_x(s)}{ds}\big|_{s=1} = -\log(\frac{L}{\gamma}) + \Psi^0(L) - \Psi^0(-\alpha)$.

Using the developments presented in [30], we have that $\widetilde{k}_1 = \widetilde{m}_1$ and the empirical expression for the first log-cumulant estimator for n samples z_i is $\widehat{\widetilde{k}}_1 = n^{-1} \sum_{i=1}^n \log z_i$. Therefore

$$\tilde{k}_1 = -\log \frac{L}{\gamma} + \Psi^0(L) - \Psi^0(-\alpha). \quad (6)$$
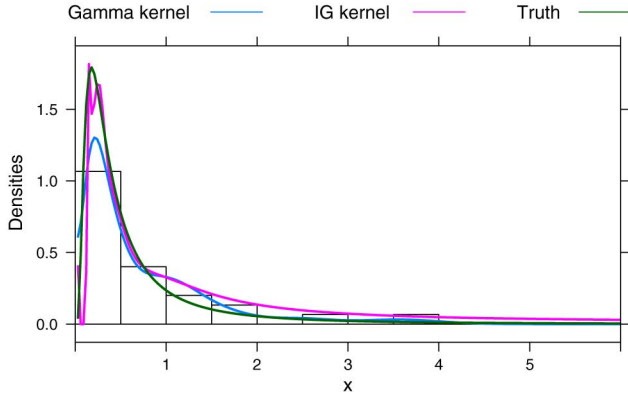
Fig. 1. Fitting the $\mathcal{G}_I^0$ density of 30 $\mathcal{G}_I^0(-3, 2, 1)$ observations in different ways, along with the histogram.
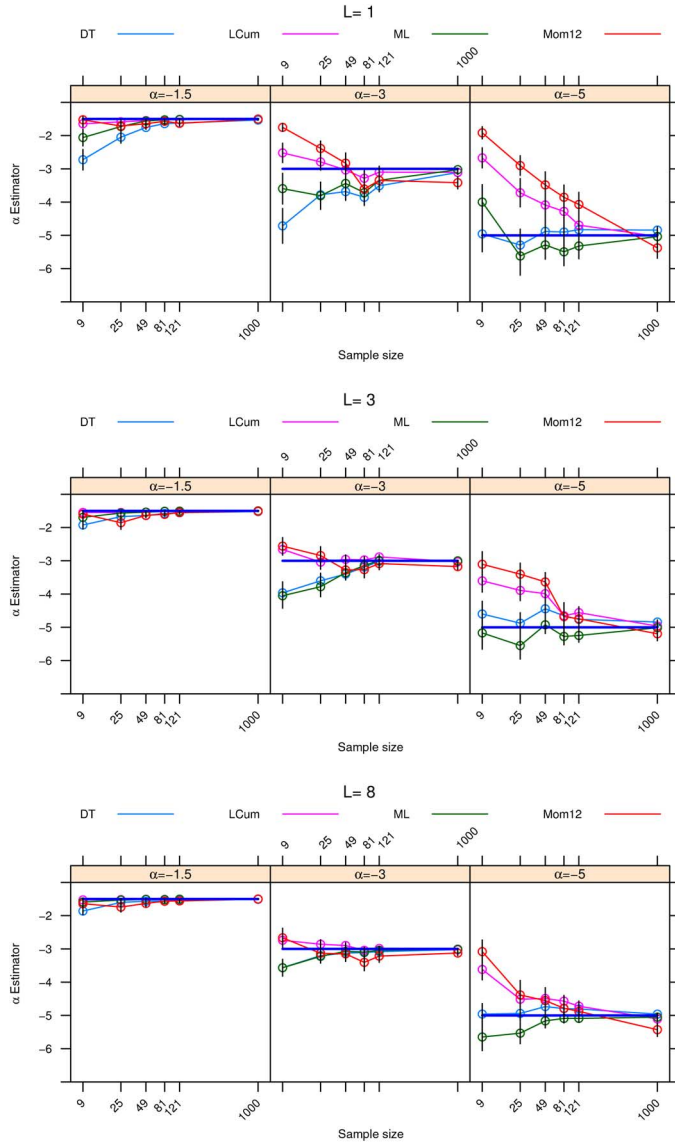

Fig. 2. Sample mean of estimates under uncontaminated data. The blue line is the true parameter.
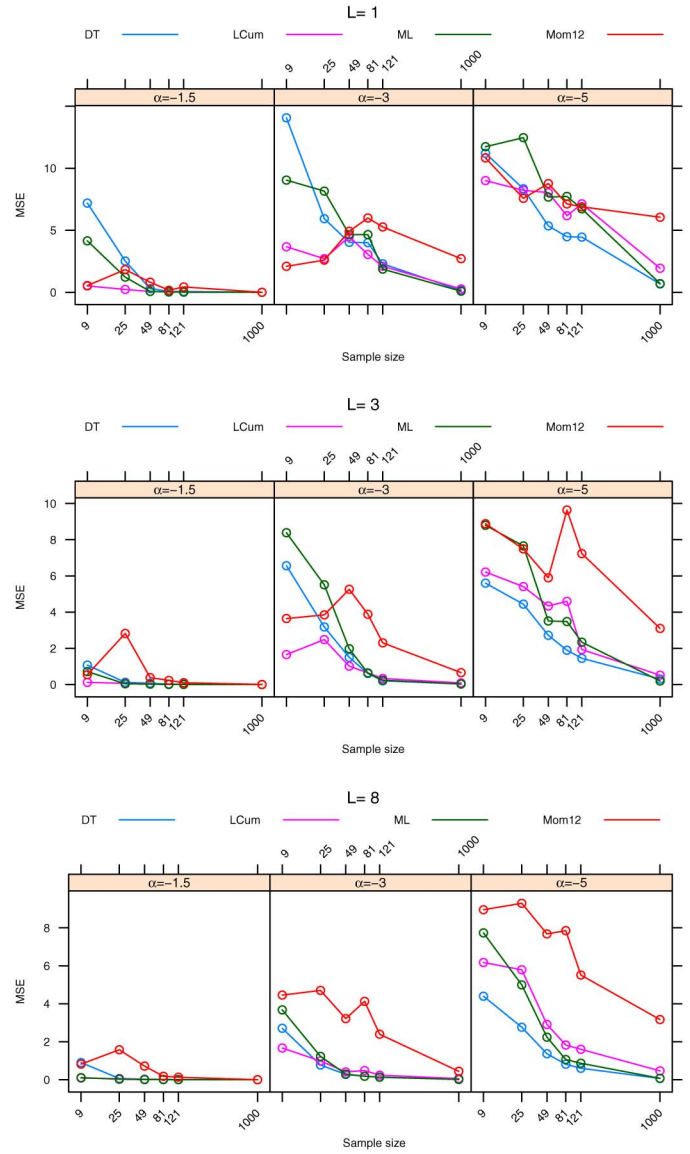

Fig. 3. Sample mean-squared error of estimates under uncontaminated data.

Assuming $\gamma^* = -\alpha - 1$, the log-cumulant estimator of α, denoted by $\widehat{\alpha}_{\text{LCum}}$ is then the solution of $\widehat{\widetilde{k}}_1 = -\log \frac{L}{1 - \widehat{\alpha}_{\text{LCum}}} + \Psi^0(L) - \Psi^0(-\widehat{\alpha}_{\text{LCum}})$, i.e., the solution of

$$\frac{1}{n} \sum_{i=1}^{n} \log z_i = -\log \frac{L}{1 - \widehat{\alpha}_{\text{LCum}}} + \Psi^0(L) + \Psi^0(-\widehat{\alpha}_{\text{LCum}}). \tag{7}$$

The ability of these estimators to resist outliers has not still been assessed.

In the following, we provide new results regarding the heavy-tailedness of the $\mathcal{G}_I^0$ distribution. This partly explains the numerical issues faced when seeking for estimators for its parameters: the distribution is prone to producing extreme values. The main concepts are taken from [31]–[33].

TABLE I
MEAN TIMES FOR SIMULATED DATA WITHOUT CONTAMINATION, $L = 1$, $n = 81$

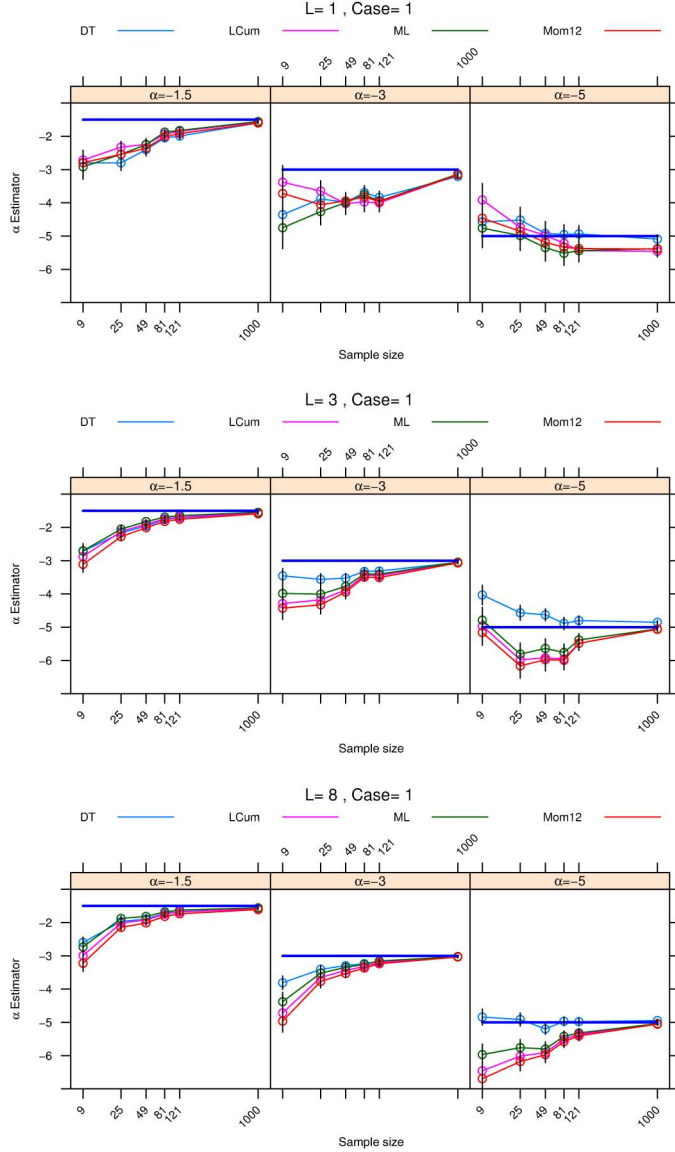| MV	DT	Moment $\frac{1}{2}$	Log-cumulant
0.003	2.223	0.0001	0.003

Fig. 4. Sample mean of estimates, Case 1 with $\alpha_2 = -15$ and $\epsilon = 0.01$.

Definition 1: The function $\ell \colon \mathbb{R} \to \mathbb{R}$ is slow-varying in the infinite, if for every $t > 0$ holds that

$$\lim_{x \to +\infty} \frac{\ell(tx)}{\ell(x)} = 1.$$

Definition 2: A probability density function $f(x)$ has heavy tails, if for any $\eta > 0$ holds that
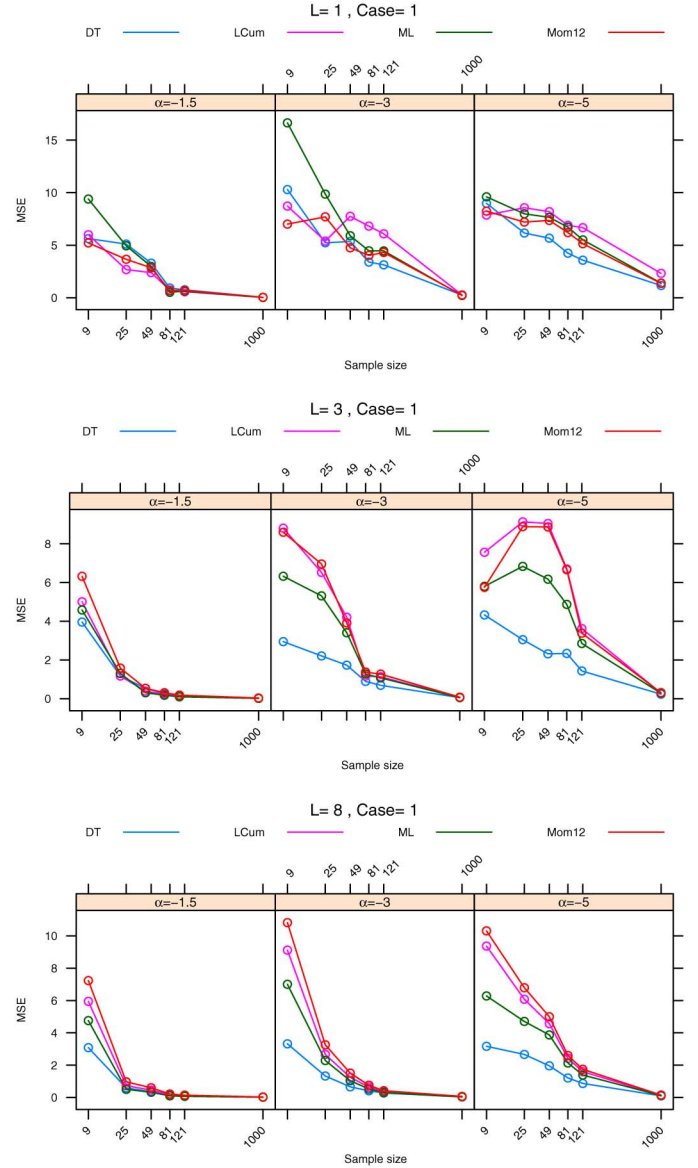
$$f(x) = \ell(x)x^{-\eta}$$

Fig. 5. Sample mean-squared error of estimates, Case 1 with $\alpha_2 = -15$ and $\epsilon = 0.01$.

where ℓ is a slow-varying function in the infinite and η is the tail index.

The smaller the tail index is, the distribution is more prone to producing extreme observations.

Proposition 1: The $\mathcal{G}_I^0$ distribution has heavy tails with tail index $1 - \alpha$.

Proof: Defining $\ell(x) = f_{\mathcal{G}_I^0}(x)x^{-\alpha+1}$, we have that

$$\lim_{x \to +\infty} \frac{\ell(tx)}{\ell(x)} = \lim_{x \to +\infty} \frac{(tx)^{L-1}(Ltx+\gamma)^{\alpha-L}(tx)^{1-\alpha}}{x^{L-1}(Lx+\gamma)^{\alpha-L}x^{1-\alpha}}$$

$$= \lim_{x \to +\infty} t^{L-\alpha}\left(\frac{Lx+\gamma}{Ltx+\gamma}\right)^{L-\alpha} = 1.$$

This holds for every $t > 0$, so ℓ is as in Definition 1. ∎

As expected, the tail index is a decreasing function on α, then the $\mathcal{G}_I^0$ distribution is more prone to producing extreme observations when the roughness parameter is bigger.

Fig. 6. Sample mean of estimates, Case 2 with $C = 100$, $\epsilon = 0.001$.
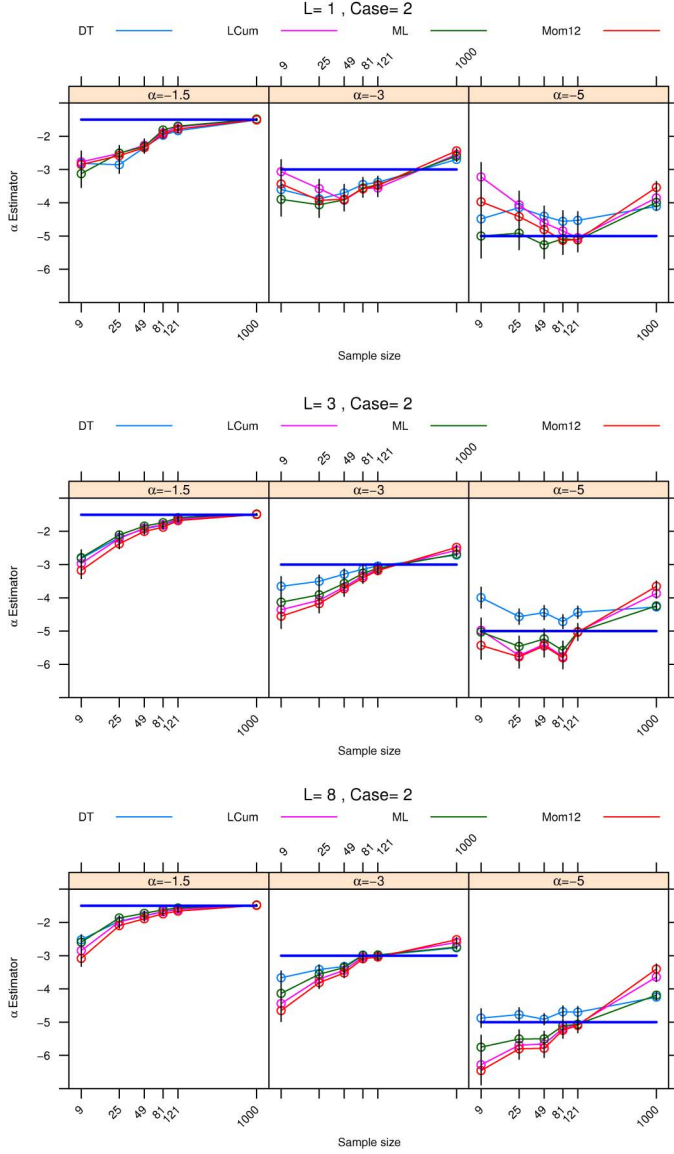
Fig. 7. Sample mean-squared error of estimates, Case 2 with $C = 100$, $\epsilon = 0.001$.

The remainder of this section is devoted to proving that the $\mathcal{G}_I^0$ model is outlier-prone. Consider the random variables $z_1, \ldots, z_n$ and the corresponding order statistics, $Z_{1:n} \leq \cdots \leq Z_{n:n}$.
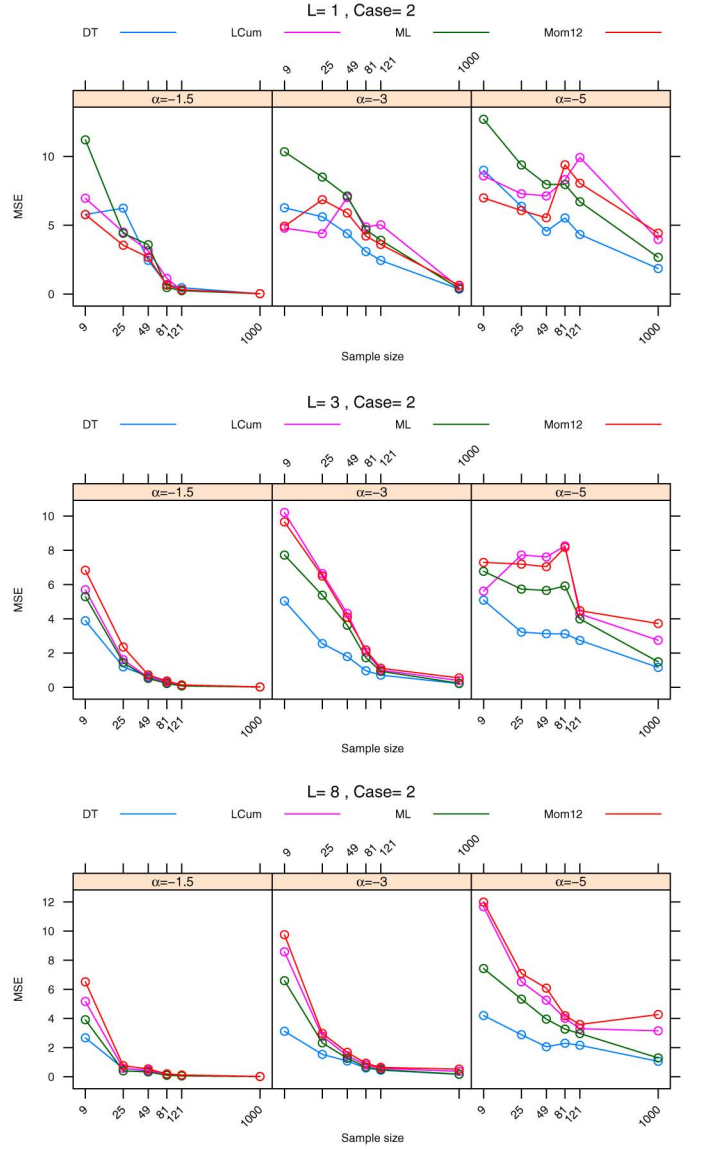
Definition 3: The distribution is absolutely outlier-prone if there are positive constants ε, δ and an integer n_0 such that

$$\Pr(Z_{n:n} - Z_{n-1:n} > \varepsilon) \geq \delta$$

holds for every integer $n \geq n_0$.

A sufficient condition for being absolutely outlier-prone is that there are positive constants ε, δ, x_0 such that the density of the distribution satisfies, for every $x \geq x_0$, that

$$\frac{f(x+\varepsilon)}{f(x)} \geq \delta. \qquad (8)$$

Since

$$\lim_{x \to +\infty} \frac{f_{\mathcal{G}_I^0}(x+\varepsilon)}{f_{\mathcal{G}_I^0}(x)} = \lim_{x \to +\infty} \left(\frac{x+\varepsilon}{x}\right)^{L-1} \left(\frac{Lx+\gamma}{L(x+\varepsilon)+\gamma}\right)^{L-\alpha}$$
$$= 1$$

we are in the presence of an absolutely outlier-prone model.

III. ESTIMATION BY THE MINIMIZATION OF STOCHASTIC DISTANCES

Information theory provides divergences for comparing two distributions; in particular, we are interested in distances between densities. Our proposal consists of, given the sample z, computing $\widehat{\alpha}$, estimator of α, as the point which minimizes the distance between the density $f_{\mathcal{G}_I^0}$ and an estimate of the underlying density function.

Cassetti *et al.* [22] assessed the Hellinger, Bhattacharyya, Rényi, and Triangular distances, and they concluded that the
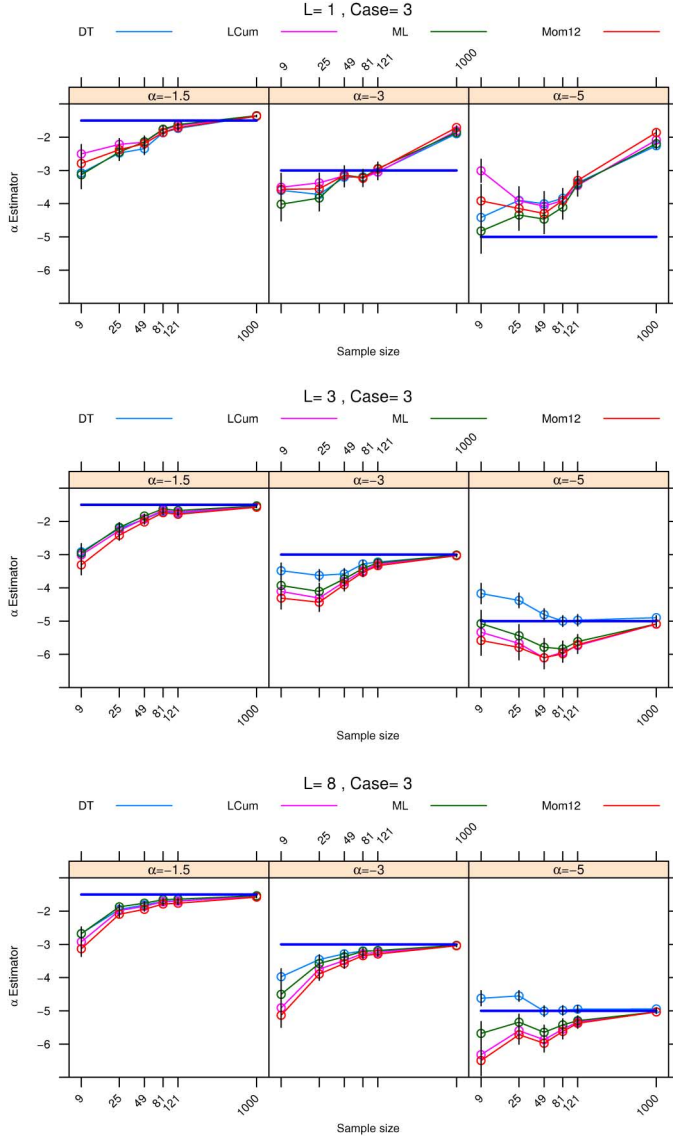
Fig. 8. Sample mean of estimates, Case 3 with $\epsilon = 0.005$ and $k = 2$.

Fig. 9. Sample mean-squared error of estimates, Case 3 with $\epsilon = 0.005$ and $k = 2$.

latter outperforms the other ones in a variety of situations. The Triangular distance between the densities f_V and f_W with common support S is given by

$$d_T(f_V, f_W) = \int_S \frac{(f_V - f_W)^2}{f_V + f_W}. \tag{9}$$

Let $\boldsymbol{z} = (z_1, \ldots, z_n)$ be a random sample of n independent $\mathcal{G}_I^0(\alpha_0, \gamma_0^*, L_0)$-distributed observations. An estimate of the underlying density function of $\boldsymbol{z}$, denoted $\widehat{f}$, is used to define the objective function to be minimized as a function of α. The estimator for the α parameter based on the minimization of the Triangular distance between an estimate of the underlying density $\widehat{f}$ and the model $f_{\mathcal{G}_I^0}$, denoted by $\widehat{\alpha}_T$, is given by

$$\widehat{\alpha}_T = \arg \min_{-20 \leq \alpha \leq -1} d_T\big(f_{\mathcal{G}_I^0}(\alpha, \gamma_0^*, L_0), \widehat{f}(\boldsymbol{z})\big) \tag{10}$$

where γ_0^* and L_0 are known and d_T is given in (9). Solving (10) requires two steps: 1) the integration of (9) using the $\mathcal{G}_I^0$ density
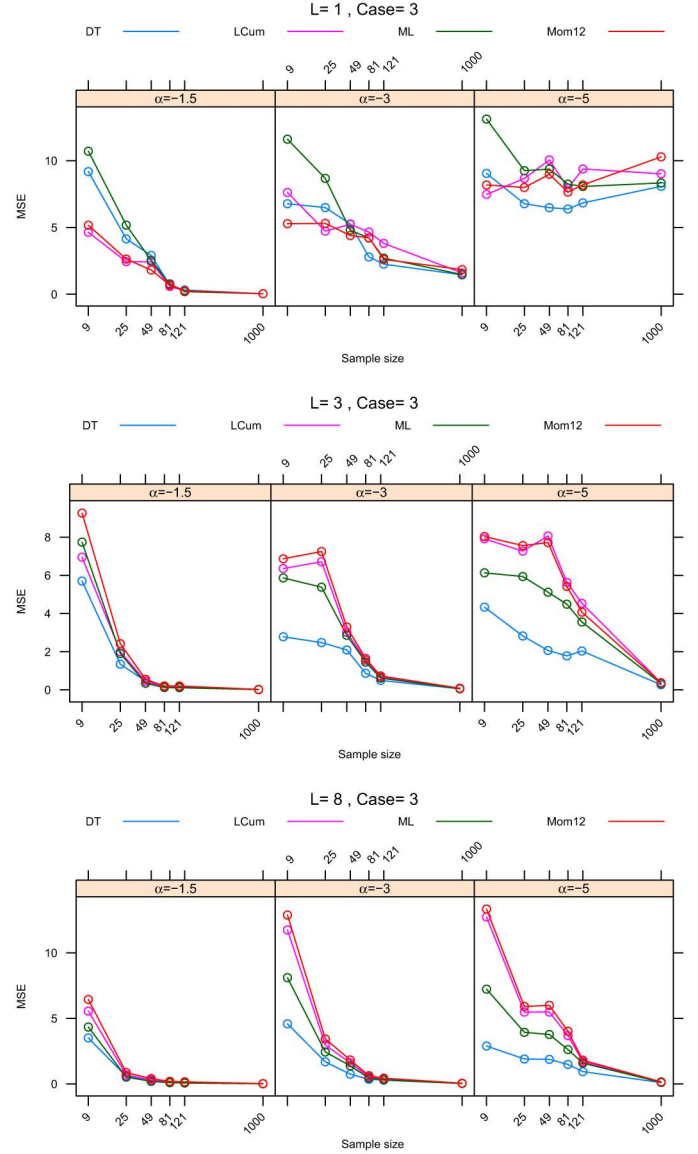
and the density estimate $\widehat{f}$, and 2) optimizing with respect to α. To the best of the authors' knowledge, there are no explicit analytic results for either problem, so we rely on numerical procedures. The range of the search is established to avoid numerical instabilities.

We also calculate numerically the ML, the $\frac{1}{2}$-moment, and the log-cumulants-based estimators using the same established range of the search and compare all the methods through the bias and the mean-squared error by simulation.

A. Estimate of the Underlying Distribution

The choice of the way in which the underlying density function is computed is very important in our proposal. In this section, we describe several possibilities for computing it, and we justify our choice.

Histograms are the simplest empirical densities, but they lack unicity and they are not smooth functions.

Given that the model we are interested is asymmetric, another possibility is to use asymmetric kernels [24], [26]. Let $\boldsymbol{z} = (z_1, \ldots, z_n)$ be a random sample of size n, with an unknown density probability function f, an estimate of its density function using kernels is given by

$$\widehat{f}_b(t; \boldsymbol{z}) = \frac{1}{n} \sum_{i=1}^{n} K(t; z_i, b)$$

where b is the bandwith of the kernel K.

Among the many available asymmetric kernels, we worked with two: the densities of the Gamma and inverse Gaussian distributions. These kernels are given by

$$K_\Gamma(t; z_i, b) = \frac{t^{z_i/b} \exp\{-t/b\}}{b^{z_i/b+1} \Gamma(z_i/b + 1)} \quad \text{and}$$

$$K_{\text{IG}}(t; z_i, b) = \frac{1}{\sqrt{2\pi b t^3}} \exp\left\{-\frac{1}{2bz_i}\left(\frac{t}{z_i} + \frac{z_i}{t} - 2\right)\right\}$$

respectively, for every $t > 0$. Empirical studies led us to employ $b = n^{-1/2}/5$.

As an example, Fig. 1 shows the $\mathcal{G}_I^0(-3, 2, 1)$ density and three estimates of the underlying density function obtained with $n = 30$ samples: those produced by the Gamma and the inverse Gaussian kernels and the histogram computed with the Freedman–Diaconis method.

As it can be seen, fitting the underlying density function using kernels is better than using a histogram. After extensive numerical studies, we opted for the K_{IG} kernel due to its low computational cost, its ability to describe observations with large variance, and its good numerical stability. In agreement with what Bouezmarni and Scaillet [26] reported, the Gamma kernel is also susceptible to numerical instabilities.

B. Contamination

Estimators in signal and image processing are often used in a wide variety of situations, so their robustness is of highest importance. Robustness is the ability to perform well when the data obey the assumed model and to not provide completely useless results when the observations do not follow it exactly. Robustness, in this sense, is essential when designing image processing and analysis applications. Filters, e.g., employ estimators based, typically, in small samples, which, more often than not, receive samples from more than one class. Supervised classification relies on samples, and the effect of contamination (often referred to as "training errors") on the results has been attested, among other works in [34].

In order to assess the robustness of the estimators, we propose three contamination models able to describe realistic departures from the hypothetical "independent identically distributed sample" assumption.

One of the sources of contamination in SAR imagery is the phenomenon of double bounce, which results in some pixels having a high return value. The presence of such outliers may provoke big errors in the estimation.

In order to assess the robustness of the proposal, we generate contaminated random samples using three types (cases) of contamination, with $0 < \epsilon \ll 1$ the proportion of contamination.

TABLE II
PERCENTAGE OF SITUATIONS FOR WHICH NO CONVERGENCE WAS OBSERVED FOR THE MOMENTS AND LOG-CUMULANT METHODS IN CASE 1, $\alpha_2 = -15$, $\epsilon = 0.01$

L	$\frac{1}{2}$-Moment	Log-cumulant
1	22.87	21.56
3	11.71	11.87
8	5.81	6.04

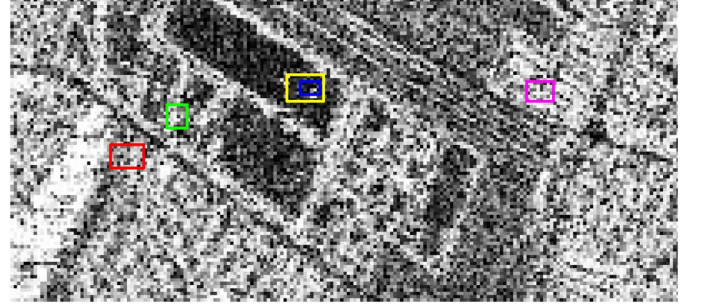
Fig. 10. Real SAR image and the regions used to estimate the α-parameter.

Let the Bernoulli random variable B with probability of success ϵ model the occurrence of contamination. Let $C \in \mathbb{R}_+$ be a large value.

1) Case 1: Let W and U be such that $W \sim \mathcal{G}_I^0(\alpha_1, \gamma_1^*, L)$, and $U \sim \mathcal{G}_I^0(\alpha_2, \gamma_2^*, L)$. Define $Z = BU + (1 - B)W$, then we generate $\{z_1, \ldots, z_n\}$ identically distributed random variables with cumulative distribution function

$$(1 - \epsilon)\mathcal{F}_{\mathcal{G}_I^0(\alpha_1, \gamma_1^*, L)}(z) + \epsilon \mathcal{F}_{\mathcal{G}_I^0(\alpha_2, \gamma_2^*, L)}(z)$$

where $\mathcal{F}_{\mathcal{G}_I^0}(\alpha, \gamma, L)$ is the cumulative distribution function of a $\mathcal{G}_I^0(\alpha, \gamma, L)$ random variable.

2) Case 2: Consider $W \sim \mathcal{G}_I^0(\alpha_1, \gamma_1^*, L)$; return $Z = BC + (1 - B)W$.

3) Case 3: Consider $W \sim \mathcal{G}_I^0(\alpha, \gamma^*, L)$ and $U \sim \mathcal{G}_I^0(\alpha, 10^k \gamma^*, L)$ with $k \in \mathbb{N}$. Return $Z = BU + (1 - B)W$, then $\{z_1, \ldots, z_n\}$ are identically distributed random variables with cumulative distribution function

$$(1 - \epsilon)\mathcal{F}_{\mathcal{G}_I^0(\alpha, \gamma^*, L)}(z) + \epsilon \mathcal{F}_{\mathcal{G}_I^0(\alpha, 10^k \gamma^*, L)}(z).$$

All these models consider departures from the hypothesized distribution $\mathcal{G}_I^0(\alpha, \gamma^*, L)$. The first type of contamination assumes that, with probability ϵ, instead of observing outcomes from the "right" model, a sample from a different one will be observed; notice that the outlier may be close to the other observations. The second type returns a fixed and typically large value C with probability ϵ. The third type is a particular case of the first, where the contamination assumes the form of a distribution whose scale is k orders of magnitude larger than the hypothesized model. We use the three cases of contamination models in our assessment.

IV. RESULTS

A Monte Carlo experiment was set to assess the performance of each estimation procedure. The parameter space

TABLE III
ESTIMATIONS OF THE α-PARAMETER USING THE SAMPLES SHOWN IN FIG. 10

Color	Size	$\widehat{\alpha}_{MV}$	$\widehat{\alpha}_T$	$\widehat{\alpha}_{Mom12}$	$\widehat{\alpha}_{LCum}$	Time MV	Time DT	Time $\frac{1}{2}$Mom	Time LCum
Magenta	100	-1.9	-2.7	-1.9	-1.7	0.03	5.85	0.03	0.02
Green	48	-2.5	-2.5	-2.9	-3.1	0.00	3.31	0.00	0.00
Blue	25	-4.9	-3.0	NA	NA	0.00	2.08	0.00	0.00
Yellow	90	-6.2	-5.1	-6.6	-6.8	0.00	5.16	0.00	0.00
Red	64	-1.8	-1.9	-1.9	-1.8	0.00	4.17	0.00	0.00

TABLE IV
SAMPLE p VALUES OF THE KOLMOGOROV–SMIRNOV TEST WITH
SAMPLES FROM THE IMAGE IN FIG. 10

| Color	p Value			
	TestMV	TestDT	TestMom	TestLCum
Magenta	0.46	0.58	0.28	0.69
Green	0.37	0.85	0.37	0.37
Blue	0.15	0.07	NA	NA
Yellow	0.63	0.98	0.76	0.22
Red	0.30	0.30	0.21	0.99

consists of the grid formed by 1) three values of roughness: $\alpha = \{-1.5, -3, -5\}$, which are representative of areas with extreme and moderate texture; 2) three usual levels of signal-to-noise processing, through $L = \{1, 3, 8\}$; 3) sample sizes $n = \{9, 25, 49, 81, 121, 1000\}$, related to squared windows of side 3, 5, 7, 9, and 11, and to a large sample; 4) each of the three cases of contamination, with $\epsilon = \{0.001, 0.005, 0.01\}$, $\alpha_2 = \{-4, -15\}$, $C = 100$, and $k = 2$.

One thousand samples were drawn for each point of the parameter space, producing $\{\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{1000}\}$ estimates of each kind. Estimates of the mean $\overline{\overline{\alpha}} = (1000)^{-1}\sum_{i=1}^{1000}\widehat{\alpha}_i$, bias $\widehat{B}(\widehat{\alpha}) = \overline{\widehat{\alpha}_i} - \alpha$, and mean-squared error $\widehat{\text{mse}} = (1000)^{-1}\sum_{i=1}^{1000}(\widehat{\alpha}_i - \alpha)^2$ were then computed and compared.

In the following figures, "ML," "T," "Mom12," and "LCum" denote the estimator based on the ML, Triangular distance, $\frac{1}{2}$-moment, and log-cumulant, respectively. Sample sizes are in the abscissas, which are presented in logarithmic scale. The estimates of the mean are presented with error bars that show the Gaussian confidence interval at the 95% level of confidence. Only a few points of this parameter space are presented for brevity, but the remaining results are consistent with what is discussed here.

Fig. 2 shows the mean of the estimators $\widehat{\alpha}$ in uncontaminated data ($\epsilon = 0$) with different values of n and L. Only two of the four estimators lie, in mean, very close to the true value: $\widehat{\alpha}_{MV}$ and $\widehat{\alpha}_T$; it is noticeable how far from the true value lie $\widehat{\alpha}_{LCum}$ and $\widehat{\alpha}_{Mom12}$ when $\alpha = -5$. It is noticeable that $\widehat{\alpha}_{ML}$ has a systematical tendency to underestimate the true value of α. Vasconcellos et al. [4] computed a first order approximation of such bias for a closely related model, and our results are in agreement with those. The estimator based on the Triangular distance $\widehat{\alpha}_T$ compensates this bias.

Fig. 3 shows the sample mean-squared error of the estimates under the same situation, i.e., uncontaminated data. In most cases, all estimators have very similar mean-squared error, not being possible to say that one is systematically the best one. This is encouraging, since it provides evidence that $\widehat{\alpha}_T$ exhibits
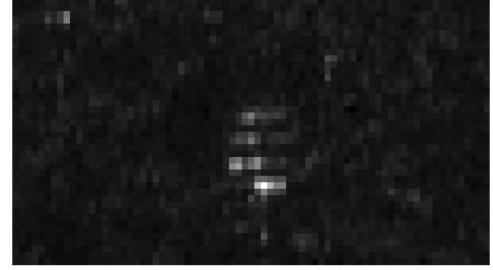
(a)
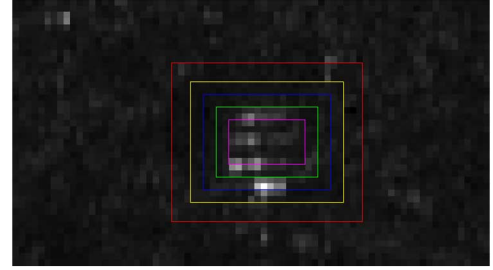
(b)

Fig. 11. Samples of several sizes in a real SAR image with a corner reflector, used to estimate the α-parameter. (a) Single-look E-SAR image with a corner reflector. (b) Regions of interest.

the first property of a good robust estimator, i.e., not being unacceptable under the true model.

The mean times of processing, measured in seconds, for each method and each case, were computed; as an example, the mean times of processing for $L = 1$ and $n = 81$ are presented in Table I. It can be seen that the new method has a higher computational cost. The other cases are consistent with this table. The details of the computer platform are presented in the Appendix.

Figs. 4 and 5 show, respectively, the sample mean and mean-squared error of the estimates under Case 1 contamination with $\alpha_2 = -15$, $\epsilon = 0.01$, and varying n and L. This type of contamination injects, with probability $\epsilon = 0.01$, observations with almost no texture in the sample under analysis. As expected, the influence of such perturbation is more noticeable in those situations, where the underlying model is further away from the contamination, i.e., for larger values of α. This is particularly clear in Fig. 5, which shows that the mean-squared errors of $\widehat{\alpha}_{ML}$, $\widehat{\alpha}_{Mom12}$, and $\widehat{\alpha}_{LCum}$ are larger than that of $\widehat{\alpha}_T$ for $L = 3, 8$, with not a clear distinction for $L = 1$ except that $\widehat{\alpha}_T$ is at least very competitive in the $\alpha = -3, -5$ cases.

Figs. 6 and 7 present, respectively, the sample mean and sample mean-squared error of the estimates under Case 2 contamination with $\epsilon = 0.001$ and $C = 100$. This type of contamination injects a constant value ($C = 100$) with probability $\epsilon = 0.001$

TABLE V
ESTIMATIONS OF THE α-PARAMETER USING THE SAMPLES SHOWN IN FIG. 11(B)

Color	Size	$\widehat{\alpha}_{MV}$	$\widehat{\alpha}_{T}$	$\widehat{\alpha}_{Mom12}$	$\widehat{\alpha}_{LCum}$	TimeMV	TimeDT	TimeMom12	TimeLogcum12
Magenta	15	−20.0	−4.1	NA	NA	0.03	1.95	0.03	0.03
Green	42	−9.2	−5.0	NA	NA	0.00	4.04	0.00	0.00
Blue	90	−3.5	−2.7	−4.7	−14.2	0.02	4.85	0.00	0.00
Yellow	156	−2.2	−1.8	−2.6	−3.4	0.01	8.35	0.00	0.00
Red	225	−1.9	−1.7	−2.1	−2.5	0.02	10.97	0.00	0.00

instead of each observation from the $\mathcal{G}_I^0$ distribution. Since we are considering samples with unitary mean, this is a large contamination. In this case, $\widehat{\alpha}_T$ is, in mean, closer to the true value than the other methods, and its mean-square error the smallest.

Figs. 8 and 9 show, respectively, the sample mean and mean-squared error of the estimates under Case 3 with $\epsilon = 0.005$ with $k = 2$. This kind of contamination draws, with probability $\epsilon = 0.005$, an observation from a $\mathcal{G}_I^0$ distribution with a scale one hundred times larger than that of the "correct" model. The behavior of the estimators follows the same pattern for $L = 3, 8$: $\widehat{\alpha}_T$ produces the closest estimates to the true value with reduced mean-squared error. There is no good estimator for the single-look case with this case of contamination.

As mentioned previously, the iterative algorithms that implement $\frac{1}{2}$-moment and log-cumulant estimators often fail to converge. In the Monte Carlo experiment, if $\frac{1}{2}$-moment or log-cumulant fail, we eliminate the estimators computed with the other methods in the corresponding iteration, so the amount of elements for calculating the mean $\overline{\widehat{\alpha}}$, the bias, and the mean-squared error is lower than 1000. As an example, Table II informs the number of such situations for Case 1 and $\alpha_2 = -15, \epsilon = 0.01$. These results are consistent with other situations under contamination, and they suggest that these methods are progressively more prone to fail in more heterogeneous areas.

Data from a single-look L-band HH polarization in intensity format E-SAR [35] image were used in the following. Fig. 10 shows the regions used for estimating the texture parameter. Table III shows the results of estimating the α parameter for each rectangular region, where NA means that the corresponding estimator is not available.

The Kolmogorov–Smirnov test (KS test) is applied to two samples: $\boldsymbol{x}$ from the image and $\boldsymbol{y}$ simulated. Babu and Feigelson [36] warn about the use of the same sample for parameter estimation and for performing a KS test between the data and the estimated cumulative distribution function. We then took the real sample $\boldsymbol{x}$, used to estimate the parameters with the four methods under assessment, and then samples of the same size $\boldsymbol{y}$ were drawn from the $\mathcal{G}_I^0$ law with those parameters. The KS test was then performed between $\boldsymbol{x}$ and $\boldsymbol{y}$ for the null hypothesis H_0 "both samples come from the same distribution," and the complementary alternative hypothesis.

Table IV shows the sample p values. It is observed that the null hypothesis is not rejected with a significance level of 5% in any of the cases. This result justifies the adequacy of the model for the data.

Fig. 11 shows the regions used for estimating the texture parameter under the influence of a corner reflector. Table V

TABLE VI
SAMPLE p VALUES OF THE KOLMOGOROV-SMIRNOV TEST USING THE SAMPLES SHOWN IN FIG. 11(B)

	p Value			
Color	TestMV	TestDT	TestMom	TestLogCum
Magenta	0.38	0.93	NA	NA
Green	0.11	0.93	NA	NA
Blue	0.01	0.11	0.400	0.11
Yellow	0.19	0.31	0.460	0.15
Red	0.23	0.12	0.008	0.02

shows the estimates of α for each rectangular region in the image of Fig. 11(b). The ML, $\frac{1}{2}$-moment, and log-cumulant estimators are unable to produce any estimate in small samples. They only produce sensible values when the sample is of at least 90 observations, The log-cumulant method is the one that requires the largest sample size to produce an acceptable estimate. The estimator based on the Triangular distance yields plausible values under contamination even with very small samples.

Table VI presents the p values of the KS test, applied as described previously to the samples of Fig. 11(b). Estimators based on $\frac{1}{2}$-moments and log-cumulants failed to produce estimates in two samples and, therefore, it was not possible to apply the procedure. The other samples did not fail to pass the KS test, except for the Blue sample and the ML estimator. These results lead us to conclude that the underlying model is a safe choice regardless the presence of contamination and the estimation procedure.

V. CONCLUSION

We proposed a new estimator for the texture parameter of the $\mathcal{G}_I^0$ distribution based on the minimization of a stochastic distance between the model (which is indexed by the estimator) and an estimate of the probability density function built with asymmetric kernels. We defined three models of contamination inspired in real situations in order to assess the impact of outliers in the performance of the estimators. The difficulties of estimating the textured parameter of the $\mathcal{G}_I^0$ distribution were further investigated and justified throughout new theoretical results regarding its heavy-tailedness.

Regarding the impact of the contamination on the performance of estimators, we observed the following. Only fractional moment and log-cumulant estimators fail to converge.

1) Case 1: regardless the intensity of the contamination, the bigger the number of looks, the smaller the percentage of situations for which no convergence if achieved.

2) Cases 2 and 3: the percentage of situations for which there is no convergence increases with the level of contamination and reduces with α.

In the single-case look, the proposed estimator does not present excellent results, but it never fails to converge, whereas the others are prone to produce useless output.

The new estimator presents good properties as measured by its bias and mean-squared error. It is competitive with the ML, fractional moment, and log-cumulant estimators in situations without contamination and outperforms the other techniques even in the presence of small levels of contamination.

For this reason, it would be advisable to use $\widehat{\alpha}_T$ in every situation, specially when small samples are used and/or when there is the possibility of having contaminated data. The extra computational cost incurred in using this estimator is, at most, the 20 times the required to compute $\widehat{\alpha}_{ML}$, but its advantages outnumber these extra computer cycles.

Appendix

Simulations were performed using the R language and environment for statistical computing [37] version 3.0.2. The `adaptIntegrate` function from the `cubature` package was used to perform the numerical integration required to evaluate the Triangular distance, the algorithm utilized is an adaptive multidimensional integration over hypercubes.

In order to numerically find $\widehat{\alpha}_{LCum}$, we used the function `uniroot` implemented in R to solve (7).

The computer platform is Intel(R) Core i7, with 8 GB of memory and 64 bits Windows 7. Codes and data are available upon request from the corresponding author.

References

[1] A. C. Frery, H.-J. Müller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

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

[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. Silva, F. Cribari-Neto, and A. C. Frery, "Improved likelihood inference for the roughness parameter of the GA0 distribution," *Environmetrics*, vol. 19, no. 4, pp. 347–368, 2008.

[6] F. Cribari-Neto, A. C. Frery, and M. F. Silva, "Improved estimation of clutter properties in speckled imagery," *Comput. Statist. Data Anal.*, vol. 40, no. 4, pp. 801–824, 2002.

[7] H. Allende, A. C. Frery, J. Galbiati, and L. Pizarro, "M-estimators with asymmetric influence functions: the GA0 distribution case," *J. Statist. Comput. Simul.*, vol. 76, no. 11, pp. 941–956, 2006.

[8] O. H. Bustos, M. M. Lucini, and A. C. Frery, "M-estimators of roughness and scale for GA0-modelled SAR imagery," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 1, pp. 105–114, 2002.

[9] S. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.

[10] F. Bujor, E. Trouve, L. Valet, J.-M. Nicolas, and J.-P. Rudant, "Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2073–2084, Oct. 2004.

[11] S. Khan and R. Guida, "Application of Mellin-kind statistics to polarimetric G distribution for SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3513–3528, Jun. 2014.

[12] A. C. Frery, F. Cribari-Neto, and M. O. Souza, "Analysis of minute features in speckled imagery with maximum likelihood estimation," *EURASIP J. Adv. Signal Process.*, vol. 2004, no. 16, pp. 2476–2491, 2004.

[13] D. M. Pianto and F. Cribari-Neto, "Dealing with monotone likelihood in a model for speckled data," *Comput. Statist. Data Anal.*, vol. 55, pp. 1394–1409, 2011.

[14] C. Arndt, *Information Measures: Information and Its Description in Science and Engineering*. New York, NY, USA: Springer, 2004.

[15] S. Aviyente, F. Ahmad, and M. G. Amin, "Information theoretic measures for change detection in urban sensing applications," in *Proc. IEEE Workshop Signal Process. Appl. Public Security Forensics*, 2007, pp. 1–6.

[16] B. Vemuri, M. Liu, S.-I. Amari, and F. Nielsen, "Total Bregman divergence and its applications to DTI analysis," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 475–483, Feb. 2011.

[17] A. Nascimento, R. Cintra, and A. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 373–385, Jan. 2010.

[18] W. B. Silva, C. C. Freitas, S. J. S. Sant'Anna, and A. C. Frery, "Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1263–1273, Mar. 2013.

[19] A. D. C. Nascimento, M. M. Horta, A. C. Frery, and R. J. Cintra, "Comparing edge detection methods based on stochastic entropies and distances for PolSAR imagery," *IEEE J. Sel. Topics Applied Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 648–663, Feb. 2014.

[20] R. C. P. Marques, F. N. Medeiros, and J. Santos Nobre, "SAR image segmentation based on level set approach GA0 model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2046–2057, Oct. 2012.

[21] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4394–4412, Oct. 2006.

[22] J. Cassetti, J. Gambini, and A. C. Frery, "Parameter estimation in SAR imagery using stochastic distances," in *Proc. 4th Asia-Pacific Conf. Synthetic Aperture Radar (APSAR)*, Tsukuba, Japan, 2013, pp. 573–576.

[23] B. Han, D. Comaniciu, Y. Zhu, and L. Davis, "Sequential kernel density approximation and its application to real-time visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1186–1197, Jul. 2008.

[24] S. Chen, "Probability density functions estimation using Gamma kernels," *Ann. Inst. Statist. Math.*, vol. 52, no. 3, pp. 471–480, 2000.

[25] O. Scaillet, "Density estimation using inverse and reciprocal inverse Gaussian kernels," *J. Nonparametric Statist.*, vol. 16, no. 1–2, pp. 217–226, 2004.

[26] T. Bouezmarni and O. Scaillet, "Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data," *Econometric Theory*, vol. 21, pp. 390–412, 2005.

[27] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 1773–1784, 2003.

[28] M. E. Mejail, A. C. Frery, J. Jacobo-Berlles, and O. H. Bustos, "Approximation of distributions for SAR images: Proposal, evaluation and practical consequences," *Latin Amer. Appl. Res.*, vol. 31, pp. 83–92, 2001.

[29] J. Gambini, M. Mejail, J. J. Berlles, and A. Frery, "Accuracy of local edge detection in speckled imagery," *Statist. Comput.*, vol. 18, no. 1, pp. 15–26, 2008.

[30] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.

[31] R. F. Green, "Outlier-prone and outlier-resistant distributions," *Amer. Statist. Assoc.*, vol. 71, no. 354, pp. 502–505, 1976.

[32] J. Danielsson, B. N. Jorgensen, M. Sarma, and C. G. de Vries, "Comparing downside risk measures for heavy tailed," *Econ. Lett.*, vol. 92, no. 2, pp. 202–208, 2006.

[33] J. Rojo, "Heavy-tailed densities," *Comput. Statist.*, vol. 5, no. 10, pp. 30–40, 2013.

[34] A. C. Frery, S. Ferrero, and O. H. Bustos, "The influence of training errors, context and number of bands in the accuracy of image classification," *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1425–1440, 2009.

[35] R. Horn, "The DLR airborne SAR project E-SAR," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, 1996, pp. 1624–1628.

[36] G. J. Babu and E. D. Feigelson, "Astrostatistics: Goodness-of-fit and all that!" in *Proc. Astron. Data Anal. Softw. Syst. XV ASP Conf. Ser.*, vol. 351, 2006, pp. 127–136.

[37] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, 2013. [Online]. Available: http://www.R-project.org/

Juliana Gambini received the B.S. degree in mathematics and the Ph.D. degree in computer science both from the Universidad de Buenos Aires (UBA), Buenos Aires, Argentina, in 1996 and 2006, respectively.

Currently, she is a Professor with the Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina, and with the Universidad Nacional de Tres de Febrero, Pcia. de Buenos Aires, Argentina. Her research interests include SAR image processing, video processing, and image recognition.

María Magdalena Lucini received the B.S. and Ph.D. degrees in mathematics from the Universidad Nacional de Córdoba, Córdoba, Argentina, in 1996 and 2002, respectively.

Currently, she is a Professor with the Universidad Nacional del Nordeste, Corrientes, Argentina, and a Researcher of CONICET, Córdoba, Argentina. Her research interests include statistical image processing and modelling, hyperspectral data, SAR, and PolSAR data.

Julia Cassetti received the B.Sc. degree in mathematics and the Postgraduate Diploma in mathematical statistics, from the Universidad de Buenos Aires (UBA), Buenos Aires, Argentina, in 2011 and 2012, respectively.

Currently, she is an Assistant Professor with the Universidad Nacional de General Sarmiento, Buenos Aires, Argentina. Her research interests include SAR imaging and applied statistics.

Alejandro C. Frery (S'92–SM'03) received the B.Sc. degree in electronic and electrical engineering from the Universidad de Mendoza, Mendoza, Argentina, in 1985, the M.Sc. degree in applied mathematics (statistics) from the Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, in 1990, and the Ph.D. degree in applied computing from the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil in 1993.

Currently, he is the Leader of LaCCAN— Laboratório de Computação Científica e Análise Numérica, Universidade Federal de Alagoas, Brazil. His research interests include statistical computing and stochastic modeling.