

# An Improved Minimum-Distance Texture Estimator for Speckled Data under the $\mathcal{G}^0$ Model

Julia Cassetti · Alejandro C. Frery

Received: 29 May 2020 / Accepted: date

**Abstract** The  $\mathcal{G}^0$  distribution is an apt model for speckled data, such as SAR imagery, because it possesses the ability to characterize areas with different degrees of texture. In the monopolarized case, this distribution depends on three parameters: the texture, the scale, and the number of looks. The first one is related to the roughness of the image, so its estimation deserves special attention. This paper proposes and compares estimation methods of the texture parameter in intensity format. We treat the multilook case. The proposal is to estimate this parameter by minimizing the triangular distance between the  $\mathcal{G}^0$  density and an estimate of the underlying density function using asymmetric kernels. We assess the properties of these estimators with analytic results and simulation. We use actual images to evaluate the performance of our proposal.

**Keywords** Synthetic aperture radar · parameter estimation · kernel estimation · minimum distance estimator · texture analysis.

## 1 Introduction

Remote sensing with Synthetic Aperture Radar (SAR) data has become a vital tool for environmental studies because it produces high-resolution images of targets and landscapes.

---

This work was partially supported by CNPq and Fapeal.

Julia Cassetti  
Instituto de Desarrollo Humano, Universidad Nacional de General Sarmiento, Pcia. de Buenos Aires, Argentina.  
Tel.: (54 11) 4469-7701/7702/7703  
Fax: (54 11) 4469-7734  
E-mail: julia.cassetti@gmail.com

Alejandro C. Frery  
School of Mathematics and Statistics, Victoria University of Wellington  
PO Box 600, Wellington 6140 New Zealand  
E-mail: alejandro.frery@vuw.ac.nz

It has several advantages over optical and infrared systems due to its ability to acquire images independently of the sunlight and weather conditions. For this reason, these images have a wide range of practical applications: environmental monitoring [8, 53], and earthquake study to map the surface deformation [57], among others. Nevertheless, SAR images have the disadvantage of having speckle noise, which appears from the use of coherent illumination [59]. This kind of noise is a characteristic of technologies that employ coherent illumination, such as microwaves, sonar, laser, and ultrasound. Speckle noise makes the analysis and interpretation of this kind of image a challenging task.

The multiplicative model has been successfully used in modeling SAR images. It describes the observed data as the product of two statistically independent random variables representing, respectively, the target information and the speckle noise. In this context, the  $\mathcal{G}^0$  model has proved to describe textured and extremely textured areas better than other models [27, 34]. Three parameters index this model: one directly associated with the texture, other being a scale parameter, and the third related to the signal-to-noise ratio.

The texture is of paramount importance, as it describes the target roughness locally. In particular, it can be related to the number of elementary backscatterers [58] and, as such, its estimation is essential.

Several works have studied parameter estimation under the  $\mathcal{G}^0$  model. Among the estimation techniques available, moments (or analogy) and maximum likelihood estimators are among the most popular. Refs. [18, 52] describe attempts to improve the performance of these estimators concerning bias and mean squared error using either resampling or analytic procedures. More recently, using a connection between the Mellin's transform, several estimators based on logcumulants and logmoments have been proposed in [4, 10, 32].

Bustos et al. [11] and Allende et al. [2] studied the performance of maximum likelihood estimators under the  $\mathcal{G}^0$

model for amplitude data, showing that they critically lose robustness for moderated outliers in extremely textured areas, whereas in textureless areas this occurs for severe outliers. As expected, the lack of robustness is more pronounced for small samples. They proposed M and AM estimators for the texture with good performance under contamination, but they suffer from numerical problems, especially with small samples.

Tison et al. [51] showed that estimators based on logcumulants are well-suited to model amplitude single-look data from urban areas, where very large values are frequently observed. Nevertheless, that work overlooks the fact that estimation under the  $\mathcal{G}^0$  distribution is more sensitive to very small values rather than to extremely large ones. In this work, we explore this fact and propose suitable solutions using kernels.

A good estimator should have good behavior if the data come from the assumed model, and it should not present large deviations against small perturbation from the assumed model. In this sense, Refs. [14, 28] are recent attempts at obtaining robust estimators for the textured parameter of the  $\mathcal{G}^0$  model for intensity data.

In general, the approach is to find the parameters that minimize dissimilarity measures ( $d$ ) between the actual density function  $f_\theta$ , which depends on the  $\theta$  parameter, and an estimation of the underlying density function  $\hat{f}$ . These are the minimum distance estimators (MDE), defined as:

$$\hat{\theta} = \arg \min_{\theta \in \Omega} d(f_\theta, \hat{f}), \quad (1)$$

where  $\theta \in \Omega$  is the parameter to be estimated, and  $\Omega$  is the parameter space. In this oncoming, two elements are needed: a measure of distance and  $\hat{f}$ .

Wolfowitz [54, 55] studied this class of estimators and showed that, under general conditions, they are strongly consistent. Beran [6] proposed an MDE estimator using Hellinger's distance between a theoretical model and a nonparametric density estimator using symmetric kernels. The work showed that this estimator is asymptotically efficient for certain parametric families of densities. Parr and Schucany [39] show that, under certain conditions, the MDE estimator between the empirical distribution function and theoretical distribution function is strongly consistent. Cao et al. [12] proposed to minimize a distance between the theoretical density function and  $\hat{f}$  using the classic symmetric kernel estimator to estimate  $f$ . They showed, under specific considerations, the strong consistency of these estimators, and studied its asymptotic normality for the  $L^2$  metric.

Refs. [40, 43] presented nonparametric kernel estimators using symmetric kernels, but when the function to be estimated has bounded support, this kind of estimators may assign probability mass outside the support [49], leading to biased estimations in a neighborhood of the boundary. Some

authors aim to use asymmetric kernel estimators to solve this problem. Chen [16, 17] presented Beta and Gamma ( $\Gamma$ ) kernels. Scaillet [47] introduced the inverse Gaussian (IG) and reciprocal inverse Gaussian (RIG) kernels. Bouezmarni and Scaillet [7] proved the consistency under the  $\Gamma$ , IG and RIG kernels. Jin and Kawczak [31] proposed the Birnbaum-Saunders (BS) and Lognormal (LN) kernels. Ref. [21] also treats these kernels. These estimators vary their shape according to the observation. This feature allows obtaining different degrees of smoothing [47], and, as they do not assign weight outside the density function support, they are free of boundary bias [17].

In the last decade, concepts related to information theory have gained interest in image processing. In particular, several measures have been proposed to reflect the closeness among the models that describe samples. Especially the concept of stochastic divergence has found applications in areas as diverse as signal and image processing [5], automatic region detection in SAR imagery [36, 48], speckle filters [41], as contrast measures. Unsupervised classification strategies were applied to Polarimetric Synthetic Aperture Radar (PolSAR) images in [13].

Cassetti et al. [14] proposed an MDE estimator for the texture parameter of the  $\mathcal{G}^0$  model and assessed different stochastic distance with a nonparametric estimation of the underlying density function using histograms. Gambini et al. [28] proposed to estimate  $f$  using the IG asymmetric kernel with a bandwidth chosen empirically; they found the estimator by minimizing Eq. (1) by inspecting a grid of values of  $\alpha$ . The authors concluded that the triangular distance is the best suited for this problem, and showed that their proposal outperforms the Logcumulant estimator in terms of mean squared error, bias, convergence, and robustness. So, the results were promising, with venues for improvement which are explored in this work.

In this paper, we improve the Gambini et al. [28] proposal. We use  $\Gamma$  and LN kernels instead of the IG kernel, showing that the latter presents a larger integrated mean square error and a higher percentage of non-convergence cases than the others. Regarding bandwidth choice, we use Least Squared Cross-Validation (LSCV) [45] for bandwidth selection in place of the empirical selection. Furthermore, we use a search algorithm to find the minimum of the MDE estimator instead of a loop through a grid of  $\alpha$  values.

We compare the performance of the proposed estimator for the texture parameter of the  $\mathcal{G}^0$  model for intensity data with Maximum Likelihood (ML) and Logcumulants (LC) methods in terms of convergence, bias and mean squared error and sensitivity to contamination. We obtain excellent results, especially for small sample sizes.

The rest of the paper unfolds as follows: Section 2 recalls the main features of the  $\mathcal{G}^0$  model for intensity data. Section 3 presents the discussion of estimation with stochas-

tic distances and asymmetric kernels. Section 4 presents the main results of the Monte Carlo study. Section 5 shows an application of our proposal to actual SAR images. Section 6 concludes the article discussing promising directions for future research.

## 2 The $\mathcal{G}^0$ Model

The return in monopolarized SAR images can be modeled as the product of two independent random variables, one corresponding to the backscatter  $X$ , and other to the speckle noise  $Y$ . In this manner,  $Z = XY$  represents the return  $Z$  in each pixel under the multiplicative model.

The speckle noise in intensity format  $Y$  is modeled as a  $\Gamma$  distributed random variable with unitary mean and shape parameter  $L \geq 1$ , the number of looks, while the backscatter  $X$  is considered to obey a Reciprocal of Gamma law. These assumptions give rise to the  $\mathcal{G}^0$  distribution for the return  $Z$ . Given the mathematical tractability and expressive power of the  $\mathcal{G}^0$  model [34, 35] it represents an attractive choice for SAR data modeling.

The density function for intensity data is given by

$$f_{\mathcal{G}^0(z)} = \frac{L^L \Gamma(L - \alpha)}{\gamma^\alpha \Gamma(-\alpha) \Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}}, \quad (2)$$

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

$$E(Z^r) = \left(\frac{\gamma}{L}\right)^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)} \frac{\Gamma(L + r)}{\Gamma(L)}, \quad (3)$$

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

Mejail et al. [34] proved a relationship between  $\mathcal{G}^0$  distributions and the Fisher-Snedecor  $F$  law. With this result, using standard functions available in most computational statistics environment as, for instance, the R programming language and environment [42], the cumulative distribution function of a  $\mathcal{G}^0(\alpha, \gamma, L)$  distributed random variable  $F_{\alpha, \gamma, L}$  can be obtained as

$$F_{\alpha, \gamma, L}(z) = Y_{2L, -2\alpha}(-\alpha z / \gamma), \quad (4)$$

for every  $z > 0$ , where  $Y_{2L, -2\alpha}$  is the cumulative distribution function of a Fisher-Snedecor random variable with  $2L$  and  $-2\alpha$  degrees of freedom. This relationship is useful for obtaining the quantiles and entropy, among other features of this distribution.

It is essential to achieve accurate and robust estimates of  $\alpha$  because it is related to the target's texture. Values close to zero ( $\alpha \in (-3, -1)$ ) suggest extreme texture, as in urban areas. As the value decreases, the texture is also reduced going, usually, from  $\alpha \in (-6, -3]$  in forest zones up to  $\alpha \in (-\infty, -6)$  in textureless targets, e.g., pasture or crops. So, its estimation deserves special attention.

Maximum likelihood estimation is the classical approach for parametric estimation of a density function because it has optimal asymptotic properties in terms of bias and variance, and well-known asymptotic distributional properties. To reduce our discussion to a single parameter, we based the forthcoming analysis on the condition of unitary mean, i.e.,  $E(Z) = 1$ , which links the texture and the brightness parameters:

$$\gamma^* = -\alpha - 1. \quad (5)$$

Let  $Z_1, \dots, Z_n$  be an independent random sample of finite size  $n$  from the  $\mathcal{G}^0(\alpha, \gamma^*, L)$  distribution. A maximum likelihood estimator of  $\alpha$  for  $L$  known, denoted  $\hat{\alpha}_{ML}$ , is any value in  $(-\infty, -1)$  that maximizes the loglikelihood function:

$$\begin{aligned} \hat{\alpha}_{ML} = \arg \min_{\alpha \in (-\infty, -1)} & \log \Gamma(L - \alpha) - \alpha \log(-1 - \alpha) - \log \Gamma(-\alpha) \\ & + \frac{\alpha - L}{n} \sum_{i=1}^n \log(-1 - \alpha + LZ_i). \end{aligned} \quad (6)$$

Notice that we are interested in finite sample size cases, for which we do not have to take into account the possibility of infinite distributional variance. We limit the search space to cases with finite distributional mean by avoiding the region  $[-1, 0)$ .

We employed the L-BFGS-B version of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [33] to find the optimum of this function taking as a starting point  $\alpha_0$  the alpha moment estimator when it exists. Otherwise we consider  $\alpha_0 = -1.5$ . This algorithm is a quasi-Newton type optimization method that allows box constraints and is widely used in Computer Graphics and Scientific Computing [23]. These methods do not need the Hessian matrix but use an approximation that is updated in each iteration and, thus, they only require the function and its gradient.

Function (6) is, in general, unimodal, but, depending on the sample, it may be monotonically decreasing or flat [25]. This behavior is responsible for the failure of algorithms to converge to a global maximum.

## 3 Minimum distance estimator and asymmetric kernels

The Minimum Distance Estimator (MDE) approach is a non-parametric methodology that proposes estimators by finding the value of the parameters that make the theoretical model as close as possible to the information provided by the sample. There are two elements involved: one is the measure of distance that allows quantifying proximity, and the other is the information that comes from the sample.

Regarding the sample information, some measures link the empirical distribution function and the theoretical model, and others measure the discrepancy between a nonparametric estimate of the underlying density function and the actual

density. In general, the distribution function is absolutely continuous, so it may be convenient to consider MDE estimators in terms of the density function.

Information theory provides divergence measures to discriminate between density functions in a statistical sense. Salicrú et al. [46] proposed a family of such measures, called  $(h, \phi)$ -divergences, that includes the Kullback-Leibler divergence, the  $\phi$ -divergences presented by Csiszár [19], and the generalizations of the  $J$ - and  $R$ -divergences defined by Taneja [50], among others. These  $(h, \phi)$ -divergences can be easily turned into distances, and into statistical tests.

Cassetti et al. [14] used the notion of minimizing a stochastic distance to propose a new estimator for the  $\alpha$  parameter in the  $\mathcal{G}^0$  model. This estimator is defined as

$$\hat{\alpha}_n = \arg \min_{\alpha \in \Omega} d(f_{\mathcal{G}^0}, \hat{f}_n), \quad (7)$$

where  $\Omega$  is the parameter space,  $d$  is a dissimilarity measure or stochastic distance between the theoretical density function  $f_{\mathcal{G}^0}$  and a nonparametric estimator  $\hat{f}_n$  of the underlying density function. The authors used histograms to compute  $\hat{f}_n$ , and assessed the Hellinger, Bhattacharyya, Rényi, and Triangular distances. They chose the Triangular distance because it presented the best performance.

The Triangular distance is defined as:

$$d_T(f_V, f_W) = \int_S \frac{(f_V(x) - f_W(x))^2}{f_V(x) + f_W(x)} dx, \quad (8)$$

where  $f_V$  and  $f_W$  are two density functions with common support  $S$ .

Gambini et al. [28] estimated the underlying density function with asymmetric kernels. In particular, the authors used the IG kernel, defined in subsection 3.1. In this work, we use  $\Gamma$  and LN kernels because they outperform the IG kernel: they have less integrated mean quadratic error and fewer cases of non-convergence than the IG kernel.

### 3.1 Gamma, Lognormal and Inverse Gaussian Kernels

Classical density estimation with kernels usually uses symmetric kernels. However, when the density has bounded support, such estimators present boundary bias. This bias appears because the symmetric kernels assign weight outside the support when the estimation takes place near the boundary.

Let  $Z = Z_1, \dots, Z_n$  be a random sample of size  $n$ , from an unknown density probability function  $f$  with positive support, an estimate of the underlying density function is:

$$\hat{f}_b(z) = \frac{1}{n} \sum_{i=1}^n K_{\theta(z,b)}(Z_i),$$

where  $K$  is the asymmetric kernel,  $b$  is the bandwidth and  $\theta(z, b)$  is the parameter vector.

Among the many available asymmetric kernels, we used the  $\Gamma$  and LN kernels defined, for every  $t > 0$ , as:

$$K_{\Gamma_{\theta(z,b)}}(t) = \frac{1}{\Gamma(\frac{z}{b} + 1)b^{\frac{z}{b}+1}} t^{-z/b} \exp\{-t/b\}, \quad (9)$$

$$K_{\text{LN}_{\theta(z,b)}}(t) = \frac{1}{t\sqrt{2\pi b}} \exp\left\{-\frac{(\log t - \log z - b^2)^2}{2b^2}\right\}, \quad (10)$$

respectively for  $t, z, b > 0$ .

As mentioned in Section 3, Gambini et al. [28] used the Inverse Gaussian kernel defined as

$$K_{\text{IG}}(t; z_i, b) = \frac{1}{\sqrt{2\pi b t^3}} \exp\left\{-\frac{1}{2bz_i} \left(\frac{t}{z_i} + \frac{z_i}{t} - 2\right)\right\}. \quad (11)$$

The rationale for using these kernels is the following: the  $\mathcal{G}^0$  model embeds the  $\Gamma^{-1}$  distribution; the  $\mathcal{G}^H$  model (also used for describing SAR data; cf. [26]) considers the IG distribution to model the backscatter; the Lognormal distribution is a widely used empirical model to describe SAR data [29].

The bandwidth is a sensitive choice in nonparametric density estimation. Gambini et al. [28] chose, empirically,  $b = n^{-1/2}/5$ . In this work, we propose the Least Square Cross-Validation [56] (LSCV) method to find an appropriate value.

Figure 1 shows an estimation example of the  $\mathcal{G}^0$  model. We generated a sample from a  $\mathcal{G}^0$  distribution of size  $n = 25$ ,  $L = 8$ ,  $\alpha = -5$  and  $\gamma = 4$ . The solid black line is the actual density function. These kernels fit reasonably well the theoretical density function, even for a small sample size. The IG kernel presents a good approximation to the true density function at the center of the function, although with heavier tails. The  $\Gamma$  and LN kernels show a better fit, the latter with higher kurtosis than the other kernels.

It is essential to have measures that quantify the quality of the estimation. We used the Integrated Mean Square Error (MISE) because it measures the average behavior of the estimator on different samples. It is defined as

$$\text{MISE}(\hat{f}_b(Z_1, \dots, Z_n)) = E \left[ \int_{\mathbb{R}} (\hat{f}_b(z, Z_1, \dots, Z_n) - f(z))^2 dz \right]. \quad (12)$$

It is important to note that these kernel estimators are free of boundary bias and achieve the convergence rate  $n^{-4/5}$  for the MISE. The scale parameter is set such that the expected value of the return  $Z$  is one. This condition leads to the following relation  $\gamma_0^* = -\alpha - 1$ .

### 3.2 Logcumulant estimator

The relationship between the moments of a random variable and its characteristic function, related to the Fourier transform of the probability density function, is well-known.



**Fig. 1**  $\hat{f}_{\mathcal{G}^0}$  for  $L = 8$ ,  $\alpha = -5$ ,  $\gamma = 4$  and  $n = 25$  using  $\Gamma$ , LN, and IG kernels, and LSCV for bandwidth selection.

When the moments of order  $k$  exist, they can be calculated through the derivatives of order  $k$  of the characteristic function.

In the same sense, Nicolas et al. [37] defined the first characteristic function of the second kind as the Mellin transform of its density when it has positive support.

Gambini et al. [28] show that the LC estimator for the  $\alpha$  parameter of the  $\mathcal{G}^0$  distribution, under the relationship (5), is the solution of

$$\frac{1}{n} \sum_{i=1}^n \log z_i = -\log \frac{L}{-\hat{\alpha}_{LC} - 1} + \Psi^0(L) - \Psi^0(-\hat{\alpha}_{LC}). \quad (13)$$

The right term of the equation (13) is a monotone decreasing function of  $\alpha$ . Therefore, this equation has a single root or none. This estimator is widely used in SAR image processing. Refs [4, 10, 32] applied this methodology due to its nice properties and good performance. More recently, Nogueira et al. [38] used the LC estimator applied in an unsupervised algorithm for SAR image segmentation.

We compare our proposal with LC and ML estimators because these are a commonplace in the literature.

### 3.3 Robustness analysis – Stylized Samples

Among the desirable properties of a good estimator, resistance to contamination is very important in practical applications. That is, it can produce reasonable estimations even when a proportion of the data does not come from the assumed model. This situation is of particular importance in the case of small sample sizes, e.g., when using filters defined

on sliding windows of size  $3 \times 3$ ,  $5 \times 5$  or, for instance,  $7 \times 7$ . These windows may contain data from areas with different textures, for example, at the edges between different regions or corner reflectors (features that produce a few very large observations with respect to the background).

In image processing and analysis, it is difficult to grant the underlying hypothesis under which the techniques being employed hold, so the ability to perform well even under deviations from these assumptions is a requirement.

In order to assess the robustness of the proposed estimators, we consider different alternatives. The first one is to consider that, in a small proportion  $\varepsilon$ , the data may suffer deviations of the theoretical model. In this case we define random variables  $W \sim \mathcal{G}^0(\alpha_1, \gamma_1^*, L)$  and  $U \sim \mathcal{G}^0(\alpha_2, \gamma_2^*, L)$ , where  $\gamma_1^* = -\alpha_1 - 1$  and  $\gamma_2^* = -\alpha_2 - 1$ . We consider the Bernoulli random variable  $B$  with a probability of success  $\varepsilon$  indicating the contamination ratio. We define  $Z = BU + (1 - B)W$  by its cumulative distribution function as  $\varepsilon F_{\alpha_2, \gamma_2^*, L}(z) + (1 - \varepsilon)F_{\alpha_1, \gamma_1^*, L}(z)$ . This schema considers that, on average, a small proportion  $\varepsilon$  of the data come from the same family but with different parameters.

Sometimes a few pixels present an extremely high return value. This phenomenon is produced by a double bounce of the electromagnetic signal, or by a corner reflector, and is one of the sources of contamination in SAR imagery. Such outliers may provoke significant errors in the estimation, as we show in an application to a real image. The return in such can be modeled as  $Z = BC + (1 - B)W$ , where  $C$  is a large value, and  $B$  and  $W$  are defined as before.

A way of assessing how the estimator  $T_n(z_1, \dots, z_n)$  behaves under contamination is by fixing  $n - 1$  observations and allowing one to vary. This is the EIF – Empirical Influence Function, but it depends on the particular sample. To avoid this, Andrews et al. [3], proposed using the  $i^{th}$  quantile of the assumed distribution as typical observations:  $z_i = F^{-1}((i - 1/3)/(n + 1/3))$ ,  $1 \leq i \leq n - 1$ . This is the SEIF – Stylized Empirical Influence Function that, among others, was used by Rousseeuw and Verboven [44] for estimation with very small samples and by Allende et al. [2] for AM estimators of the texture under the amplitude  $\mathcal{G}^0$  law.

The quantiles of a  $\mathcal{G}^0(\alpha, \gamma, L)$  random variable can be obtained by  $\Upsilon_{2L, -2\alpha}^{-1}$ , the inverse of (4), also available in most statistical computing environments.

### 4 Performance Analysis

This section presents empirical results about the quality of estimators for the texture parameter under the  $\mathcal{G}^0$  distribution. We will consider that an algorithm converges if it finds an estimate for  $\alpha$  in the interval  $[-20, -1]$ .

Figure 2 shows the  $\mathcal{G}^0$  densities for  $\alpha = \{-20, -25\}$ ,  $\gamma = \gamma^*$ , and  $L = \{3, 8\}$ . These densities are very similar since

the  $\mathcal{G}^0$  law converges in distribution to the  $\Gamma(L, L)$  law when  $\alpha \rightarrow -\infty$  and  $\gamma = \gamma^*$  [24]. Thus, our approach assumes that  $\alpha$  takes values in the range of  $[-20, -1]$ .

The same happens for non-convergence cases. The IG kernel has the highest percentage of these cases compared to the  $\Gamma$  and LN kernel.



(a)  $L = 3$



(b)  $L = 8$

**Fig. 2**  $\mathcal{G}^0(\alpha, \gamma^*, L)$  densities.

Many methods of image filtering and edge detection use sliding masks to estimate parameters. These mask are usually of size  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$  and  $11 \times 11$ . For this reason, we chose  $n = 9, 25, 49, 81$  and  $121$  as the sample sizes to perform the empirical analysis.

First of all, we evaluate the performance of the  $\Gamma$ , LN, and IG kernels studying MISE and the percentage of non-convergence cases, in order to choose the kernel that produces the best fits of the data. We use the LSCV method to find the bandwidth, and, for IG kernel, we also review the empirical bandwidth ( $\widehat{IG}_E$ ) studied in Ref. [28].

Table 1 shows the values of  $\widehat{MISE}$  and non-convergence percent. It can be seen that IG and  $\widehat{IG}_E$  present MISE values of several orders of magnitude higher than the other kernels.

**Table 1** Estimated MISE and percentage of non-convergence cases for  $L = 3$ .

$\alpha$	$n$	$\widehat{MISE}$				% non convergence cases			
		$\Gamma$	LN	IG	$\widehat{IG}_E$	$\widehat{\alpha}_\Gamma$	$\widehat{\alpha}_{LN}$	$\widehat{\alpha}_{IG}$	$\widehat{\alpha}_{IG_E}$
-1.5	9	0.41	0.81	6.12	41.33	0	0.4	0.4	0
	25	0.12	0.18	2.46	17.02	0	0	0	0
	49	0.08	0.54	1.16	9.67	0	0	0	0
	81	0.06	0.08	0.75	6.13	0	0	0	0
	121	0.06	0.08	0.53	4.00	0	0	0	0
-3	9	0.25	0.56	12.48	50.02	5.2	7.2	8	6.8
	25	0.08	0.11	3.22	25.40	0.2	1	0.6	0.6
	49	0.04	0.07	0.69	16.40	0	0.4	0	0
	81	0.03	0.03	0.25	11.58	0	0	0	0
	121	0.02	0.03	0.17	8.98	0	0	0	0
-5	9	0.24	0.43	15.32	63.38	9.6	13.2	17	13.2
	25	0.07	0.09	4.65	32.94	3.4	5	2	1.6
	49	0.04	0.06	0.68	24.41	1.4	1.2	1	0.4
	81	0.03	0.03	0.29	18.91	0.4	0.8	0.2	0
	121	0.02	0.02	0.18	15.92	0	0.2	0	0
-8	9	0.24	0.38	18.11	73.08	16.6	19	26.2	18
	25	0.07	0.11	5.13	40.48	81.8	11	8.8	4.2
	49	0.04	0.05	0.93	30.23	5	5.4	4.2	1.2
	81	0.03	0.03	0.33	25.71	4.4	4	2	0.2
	121	0.02	0.02	0.21	21.19	1.8	2	0.8	0.4

Therefore, we chose  $\Gamma$  and LN kernels to estimate the underlying density function.

The following analysis was performed through a Monte Carlo experiment, which consists of 500 independent replications for each of several parameter values:  $L \in \{3, 8\}$  to consider multilook case,  $\alpha \in \{-1.5, -3, -5, -8\}$  to represent different levels of texture, and  $n \in \{9, 25, 49, 81, 121, 500\}$  to consider different scenarios of window sizes, and a large sample situation.

Although the variance of a  $\mathcal{G}^0$ -distributed random variable is infinite when  $\alpha > -2$ , cf. (3), we need to assess the behavior of estimators in such cases. As noted by, among others Ref. [15, Fig. 5], one should expect that extremely heterogeneous samples as, for instance, those from urban areas, are described in this region of the parameter space.

Each replication produces estimates  $\{\widehat{\alpha}_1, \dots, \widehat{\alpha}_{500}\}$  with which we compute the sample mean  $\widehat{\alpha} = (500)^{-1} \sum_{i=1}^{500} \widehat{\alpha}_i$ , the sample bias  $\widehat{B}(\widehat{\alpha}) = \widehat{\alpha} - \alpha$ , and the sample mean squared error  $\widehat{mse} = (500)^{-1} \sum_{i=1}^{500} (\widehat{\alpha}_i - \alpha)^2$ .

We compare four estimators for the multilook case:  $\widehat{\alpha}_{ML}$ ,  $\widehat{\alpha}_{LC}$ ,  $\widehat{\alpha}_\Gamma$  ( $\Gamma$  kernel) and  $\widehat{\alpha}_{LN}$  (LN kernel).

The performance of the proposed techniques is assessed twofold: in pure and contamination cases. We also assess the robustness of these estimators employing their Stylized

Empirical Influence Functions (SEIFs) and contaminating the data, as explained in subsection 3.3. The estimators were also evaluated in terms of the percentage of situations for which there was no convergence, and by their computational cost. For  $\hat{\alpha}_{LN}$  we also account, as a case of non-convergence, for situations where no solution was found for equation (13).

#### 4.1 Computational Information

Table 2 shows the percentage of non-convergence cases for  $L = 3$ ; the situation for  $L = 8$  yields similar values and, thus, is not reported for brevity. It can be seen that the  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  estimators have the highest values of lack of convergence followed by  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_\Gamma$ .

**Table 2** Percentage of non-convergence cases,  $L = 3$ .

$\alpha$	$n$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_\Gamma$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
-1.5	9	0	0	0.4	2.8
	25	0	0	0	0.2
-3	9	13	5.2	7.2	28.4
	25	1	0.2	1	11.4
	49	0.2	0	0.4	3.8
	81	0	0	0	2.4
	121	0	0	0	0.2
-5	9	26.8	9.6	13.2	35.2
	25	10	3.4	5	28.6
	49	3.4	1.4	1.2	18.6
	81	0.2	0.4	0.8	15.8
	121	0.4	0	0.2	9.6
-8	500	0	0	0	0.6
	9	39.6	16.6	19	44.2
	25	28.6	9	11	36.4
	49	18.4	5	5.4	31.6
	81	12	4.4	4	27.2
	121	5.8	1.8	2	24.6
	500	0	0	0.2	9

Such a lack of convergence is more pronounced for homogeneous ( $\alpha = -8$ ) and textured ( $\alpha = -5$ ) areas. This issue had been observed for amplitude data in [25] for  $\hat{\alpha}_{ML}$ , and is related to the decreasing curvature of the likelihood function.

Table 3 shows the mean system time, in seconds, for five hundred replications for each estimator, considering  $\alpha = -5$  and sample  $n = 81$ . Albeit  $\hat{\alpha}_\Gamma$  and  $\hat{\alpha}_{LN}$  require much more processing time than the others estimators, the former fail to converge about half of the times that the last.

**Table 3** Mean system time for uncontaminated data,  $L = 3$  and  $n = 81$ .

ML	$\Gamma$	LN	LC
0.003	1.87	1.87	0.003

#### 4.2 Simulation Results – Pure Cases

Figure 3 shows the bias and mean square error (MSE) for uncontaminated data,  $L = 3$ , and varying sample size. The results are shown in the semilogarithm scale. The iterations considered are those where all methods converge.

We also plot the confidence interval of approximately 95 % level for each estimate, using the percentile method described in Ref. [9]. This method consists in using the percentile  $(\alpha/2)$  and  $(1 - \alpha/2)$  of the distribution of  $\hat{\alpha}$ . This method was evaluated by [9] in the case where the underlying distribution is exponential, where the symmetry of the distribution fails. The author shows that this interval has a better performance than using the normal approximation.

It can be seen that, for a fixed  $n$  and as  $\alpha$  decreases, both  $\widehat{\text{Bias}}$  and  $\widehat{\text{MSE}}$  increase in most methods. It is important to note that  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  on the one hand, and  $\hat{\alpha}_\Gamma$  and  $\hat{\alpha}_{LN}$  on the other hand, have similar behavior in both bias and MSE in most cases. This similar behavior can also be observed in the confidence intervals:  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  produce wider intervals than the other estimators, showing that they are less accurate than MDE.

For extremely heterogeneous ( $\alpha > -3$ ) and heterogeneous areas ( $\alpha \in (-6, -3]$ ), the MDE estimators have a better performance than the rest. For moderately heterogeneous areas, all methods have similar behavior; however, MDE has a lower MSE. For homogeneous areas,  $\hat{\alpha}_{ML}$  is the one with the smallest bias, although the MSE is comparable to the MDE estimators except for moderate size samples where the MSE is the largest.

A similar analysis was done for the case  $L = 8$ . Figs. 4(a) and 4(b) show that  $\hat{\alpha}_{ML}$  has the best performance for homogeneous areas in terms of bias, while  $\hat{\alpha}_{LN}$  outperforms the rest in extremely textured areas and with moderate texture.

Fig. 5 shows estimates of the density of these estimators for  $\alpha = -3$  and  $L = 3$ . As the sample size increase, the mean of these estimators approaches the true value of the texture parameter, and their distributions become more symmetric. Moreover, from  $n = 49$  onwards the shape of the  $\hat{\alpha}_{LN}$  distribution is very similar to that of  $\hat{\alpha}_{ML}$ .

In order to investigate the asymptotic behavior of these estimators, we studied estimates of their density functions for  $n = 500$  and all the texture parameters. The result of this analysis is shown in Fig. 6, where the vertical blue lines indicate the true value. Notice that the ranges of the y-axis are different among the plots. The heavytailedness of these densities is noticeable. For  $\alpha = -1.5$ , all are quite concentrated around the true parameter value. As the texture parameter reduces, the densities more spread. This issue has been previously noted in Refs. [2, 14, 18, 25], among others. The estimator which suffers the least with the texture reduction is  $\hat{\alpha}_{ML}$ , due to its optimal asymptotic properties.



**Fig. 3** Sample Bias and MSE with uncontaminated data and  $L = 3$ .

Table 4 presents the skewness and kurtosis values for simulated data and  $L = 3$ . The skewness is negative in all the cases, except for  $\alpha = -1.5$  and  $n = 121,500$  for the  $\hat{\alpha}_{LN}$  estimator. As the sample size increases, the skewness decreases in absolute value. For textured and homogeneous areas  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$  and  $\hat{\alpha}_{LN}$  present similar asymmetry. In extremely textured areas,  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  are more symmetric than the other estimators. All the estimators present large values of kurtosis in most cases, showing heavytailedness and concentration around the mean, especially for extremely and moderately textured zones.

Table 5 shows the variance of the estimators. It can be seen the decrease in variance as the sample size increases. For textured and homogeneous areas, the MDE estimators present lower variance than the others for all sample sizes. For  $n = 500$ , all variances are comparable except for  $\hat{\alpha}_{LC}$ , which has a larger value. This suggests that the asymptotic variance of the MDE estimators accompanies that of  $\hat{\alpha}_{ML}$ .

According to the study for data without contamination, we conclude that MDE estimators are competitive against the other estimators for uncontaminated data, and that they have better performance than the rest of the methods in some of the cases evaluated; this improved performance is noticeable with the  $\hat{\alpha}_{LN}$  estimator.

#### 4.3 Simulation Results – Stylized Samples

Figures 7 and 8 show the SEIFs for  $L = 3$ ,  $n = 25$  and,  $\alpha = -1.5, -3$  and  $\alpha = -5, -8$  respectively. The red ticks

marks are the maximum and minimum theoretical quantiles, i.e.,  $F_{\alpha, \gamma^*, L}^{-1}(2/(3n+1))$  and  $F_{\alpha, \gamma^*, L}^{-1}((3n-4)/(3n+1))$ , respectively. The horizontal line is the true value.

Again, for  $\alpha = -3, -5, -8$ , we see a similar behavior between  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_{LC}$  on the one hand, and  $\hat{\alpha}_T$  and  $\hat{\alpha}_{LN}$  on the other hand. In extremely textured areas  $\hat{\alpha}_{LN}$  has the best performance, while for  $z$  values lower than the last quantile and for the rest of the textures, both  $\hat{\alpha}_T$  and  $\hat{\alpha}_{LN}$  have a better performance compared to the other estimators. Moreover,  $\hat{\alpha}_{ML}$  does not converge for moderately heterogeneous and homogeneous zones, while  $\hat{\alpha}_{LC}$  does not converge for homogeneous zones. All estimators behave similarly for large values of  $z$ . This behavior shows the sensitivity and loss of robustness for  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  versus MDEs.

Figures 9(a), 9(b), 10(a) and 10(b) show the SEIFS for  $L = 8$  and the same values of  $n$  and  $\alpha$ . The performance of the estimators for this case is similar to that described for  $L = 3$ . It should be mentioned that all methods converge for this case where there is less presence of speckle noise.

#### 4.4 Simulation Results – Contamination Data

In the following, we show the performance, throughout the bias and mean square error, of the estimators under contamination. Figures 11 and 12 present these quantities under contaminated data for  $L = 3$  and  $L = 8$  respectively. In this case we considered  $\varepsilon = 0.05$

At this level of contamination, it can be seen more clearly how the estimators are grouped by performance,  $\hat{\alpha}_{ML}$  and



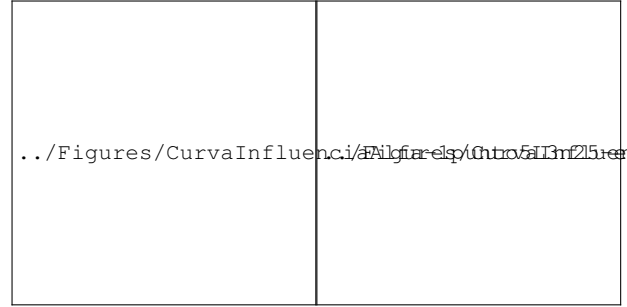


**Table 4** Skewness and Kurtosis for simulated data,  $L = 3$ 

$\alpha$	$n$	Skewness				Kurtosis			
		$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
-1.5	9	-4.86	-3.75	-6.95	-8.67	42.51	26.45	70.19	98.31
	5	-1.49	-1.91	-2.73	-5.03	6.27	8.58	20.79	47.22
	49	-0.89	-1.91	-1.07	-2.42	3.80	9.58	5.25	14.47
	81	-0.70	-2.23	-0.38	-1.55	3.42	12.21	4.16	7.76
	121	-0.64	-3.02	0.31	-1.12	3.60	16.93	6.42	5.55
	500	-0.38	-2.21	2.82	-0.51	3.16	8.41	11.47	3.54
-3	9	-2.39	-2.36	-3.42	-3.38	10.28	11.25	17.03	15.95
	25	-3.07	-4.64	-4.12	-2.99	15.68	37.20	26.88	13.94
	49	-3.87	-4.54	-4.34	-3.70	30.93	43.69	31.26	22.17
	81	-1.52	-1.84	-1.86	-3.09	6.52	9.81	9.28	15.13
	121	-2.33	-1.36	-2.17	-4.09	15.12	6.88	15.10	33.19
	500	-0.46	-0.53	-0.53	-1.29	3.29	3.31	3.33	6.21
-5	9	-1.94	-3.45	-2.64	-2.59	7.15	20.84	12.69	9.93
	25	-1.69	-3.05	-2.68	-1.93	6.04	18.47	14.18	6.41
	49	-1.82	-1.30	-2.99	-1.91	6.56	5.02	18.13	6.51
	81	-2.17	-2.51	-2.07	-1.69	10.59	15.00	9.22	5.68
	121	-2.21	-2.04	-2.01	-1.93	11.46	10.77	8.57	7.07
	500	-0.93	-0.85	-1.10	-2.13	4.69	4.62	5.11	10.37
-8	9	-1.55	-2.68	-2.69	-2.20	4.83	14.01	11.81	8.00
	25	-0.96	-2.50	-2.11	-1.67	3.29	13.58	8.68	5.12
	49	-1.12	-2.17	-1.92	-1.50	3.63	9.71	7.90	4.65
	81	-1.09	-1.82	-1.69	-1.35	3.52	8.34	7.35	4.21
	121	-1.17	-1.35	-1.71	-1.26	4.15	5.38	7.07	4.27
	500	-1.33	-1.26	-1.52	-1.35	6.21	6.11	7.39	4.50

**Table 5** Variance for simulated data,  $L = 3$ .

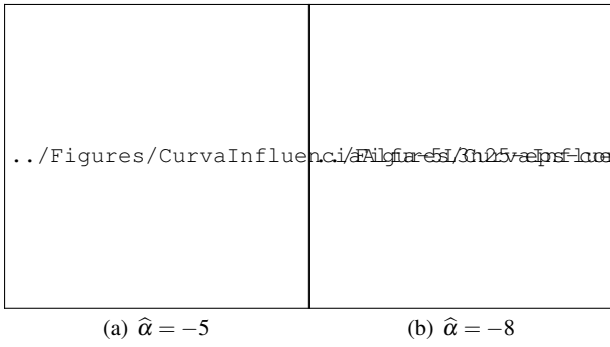
$\alpha$	$n$	Variance			
		$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
-1.5	9	0.28	0.34	0.39	1.90
	25	0.05	0.07	0.05	0.20
	49	0.02	0.03	0.02	0.04
	81	0.01	0.02	0.01	0.02
	121	0.01	0.02	0.01	0.01
	500	0.00	0.05	0.01	0.00
-3	9	5.58	2.21	5.30	8.03
	25	5.03	2.46	2.40	5.90
	49	1.62	0.85	1.31	3.93
	81	0.54	0.39	0.44	3.28
	121	0.36	0.23	0.27	1.67
	500	0.06	0.05	0.07	0.19
-5	9	10.75	3.78	5.39	12.55
	25	10.98	3.47	3.83	11.75
	49	9.49	2.31	3.87	10.65
	81	5.11	2.32	2.33	10.78
	121	2.86	1.55	1.86	8.03
	500	0.54	0.41	0.58	3.38
-8	9	16.17	3.65	6.92	11.56
	25	13.94	4.50	6.28	15.44
	49	15.18	5.05	5.75	13.45
	81	13.25	5.32	5.09	15.29
	121	10.02	4.50	5.25	12.02
	500	3.27	1.95	2.72	10.24

(a)  $\hat{\alpha} = -1.5$ (b)  $\hat{\alpha} = -3$ **Fig. 7** SEIF for  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$ ,  $\hat{\alpha}_{LN}$ ,  $\hat{\alpha}_{LC}$  para  $L = 3$ ,  $n = 25$  y  $\alpha = -1.5, -3$ .

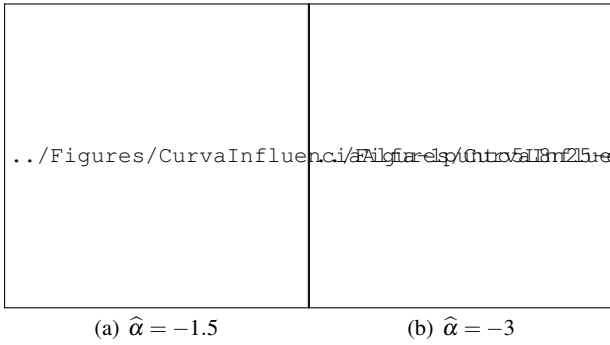
the E-SAR system [30] in L band, HH polarization, and intensity format. This image has  $300 \times 250$  pixels and mainly comprises two different growth areas.

The original image is single-look. To obtain comparable multilook data, we computed new pixels averaging the intensity values in  $2 \times 2$  non-overlapping windows. This technique is known as pyramidal processing [1], and the resulting image is multilook.

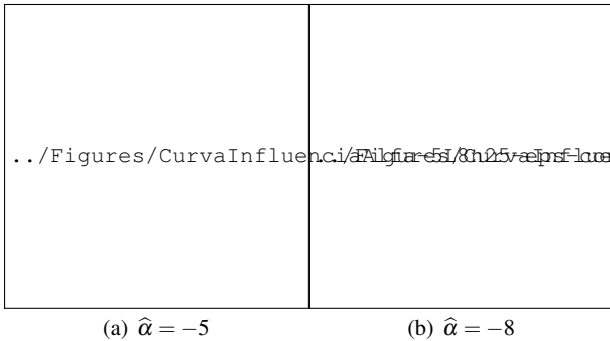
The number of looks, defined by  $E(I)^2/\text{Var}(I)$  in textureless areas, is not known, but it can be estimated by the equivalent number of looks (ENL) using uncorrelated data and the moment estimator  $\text{ENL} = 1/\widehat{\text{CV}}^2$ , the reciprocal of the sample coefficient of variation  $\widehat{\text{CV}} = \hat{\sigma}/\hat{\mu}$ , where  $\hat{\sigma}$  is



**Fig. 8** SEIF for  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$ ,  $\hat{\alpha}_{LN}$ ,  $\hat{\alpha}_{LC}$  para  $L = 3$ ,  $n = 25$  y  $\alpha = -5, -8$ .



**Fig. 9** SEIF for  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$ ,  $\hat{\alpha}_{LN}$ ,  $\hat{\alpha}_{LC}$  para  $L = 8$ ,  $n = 25$  y  $\alpha = -1.5, -3$ .



**Fig. 10** SEIF for  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$ ,  $\hat{\alpha}_{LN}$ ,  $\hat{\alpha}_{LC}$  para  $L = 8$ ,  $n = 25$  y  $\alpha = -5, -8$ .

the sample standard deviation and  $\hat{\mu}$  the sample mean. In order to find the ENL we selected manually four textureless areas and calculated  $ENL_i$  for  $i = 1, \dots, 4$ . The final ENL is their average weighted by the size of the areas. We then set  $\hat{L} = 3.21$  in the analysis of the multilook image.

We found approximate confidence intervals at the 95 % level with bootstrap and the percentile method [20], using 2000 independent replications. This method uses the percentiles  $\theta_{(\alpha)}^*$  and  $\theta_{(1-\alpha)}^*$  of the distribution of  $\hat{\theta}$  generated by the bootstrap samples, where  $\theta_{(\alpha)}^*$  and  $\theta_{(1-\alpha)}^*$  are the sample percentiles of the distribution of  $\hat{\theta}$ . According to Efron and Tibshirani [22], this method is preferable to the standard normal interval when there is evidence of non-normality. Since

our smallest sample size of interest is  $n = 9$ , and the number of permutations is  $n! = 362880$ , we obtained the bootstrap samples by uniform sampling 2000 permutations.

Figure 13 shows three samples: yellow, red, and blue. The size of the first two is 16, and the latter has a size 12. The red sample is a shift from the yellow one in a row of the image data matrix, and the blue sample is contained within the yellow sample.

Table 6 shows the estimates obtained with the four methods in each sample. It can be seen that the estimates differ considerably in the yellow sample: both  $\hat{\alpha}_{ML} = -7.21$  and  $\hat{\alpha}_{LC} = -6.74$  are compatible with a homogeneous zone, while  $\hat{\alpha}_T = -4.34$  indicates that it is a heterogeneous one, and  $\hat{\alpha}_{LN} = -3.23$  is on the limit of affirming whether the area is heterogeneous or extremely heterogeneous.

On the one hand, both  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  have substantially changed the estimations value between the yellow and the red samples, indicating a different kind of textures. On the other hand,  $\hat{\alpha}_T$  and  $\hat{\alpha}_{LN}$  changed their values, but the type of zone they describe was not modified.

Table 6 also presents the bootstrap confidence intervals for each of the cases studied. It is important to mention that we only considered cases where the algorithm for the search of the texture parameter estimate converges. The final bootstrap sample size is indicated with the variable  $n_b$  and it is mentioned as effective sample size. It can be observed that  $\hat{\alpha}_{LC}$  and  $\hat{\alpha}_{ML}$  methods yielded the smallest effective sample sizes, and that the most accurate confidence intervals were achieved by  $\hat{\alpha}_{LN}$ . These results suggest that this last method is the most dependable in term of algorithm convergence.

Then, based on the results of the simulations obtained for the SEIF corresponding to Figures 7(b) and 7(b), that show the lack of robustness of  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  when contaminated with a small value, we re-estimate this parameter removing the smallest observation that corresponds to a value 86 % lower than the average. Table 7 shows the estimates after this censoring.

Again, it can be observed a significant change in the estimated values of the  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$ , giving now estimates compatible with a homogeneous zone. This shows the robustness of the stochastic distance estimator against the  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  estimators. The table also presents the bootstrap confidence intervals and the effective sample sizes  $n_b$ . We observe the same behaviour as before:  $\hat{\alpha}_{LN}$  has the largest  $n_b$  and the shortest confidence interval.

We performed another analysis of the behavior of the estimation methods with the five samples shown in Figure 14. We chose nested squared samples of sides  $n = 9, 25, 49, 81$  and 121. Figure 15 shows the estimates  $\hat{\alpha}$  for each method and each sample as functions of the sample size  $n$ .

Table 8 shows the estimates, along with the confidence intervals and effective bootstrap sample sizes (denoted as  $n_b$ ). It can be seen that  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  have similar performance,



**Fig. 11** Bias and MSE for contaminated data,  $\varepsilon = 0.05$  and  $L = 3$ .

**Table 6** Estimates of  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_\Gamma$ ,  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_{LC}$  for three samples from the ESAR image, along with their bootstrap confidence intervals and effective bootstrap sample sizes ( $n_b$ ).

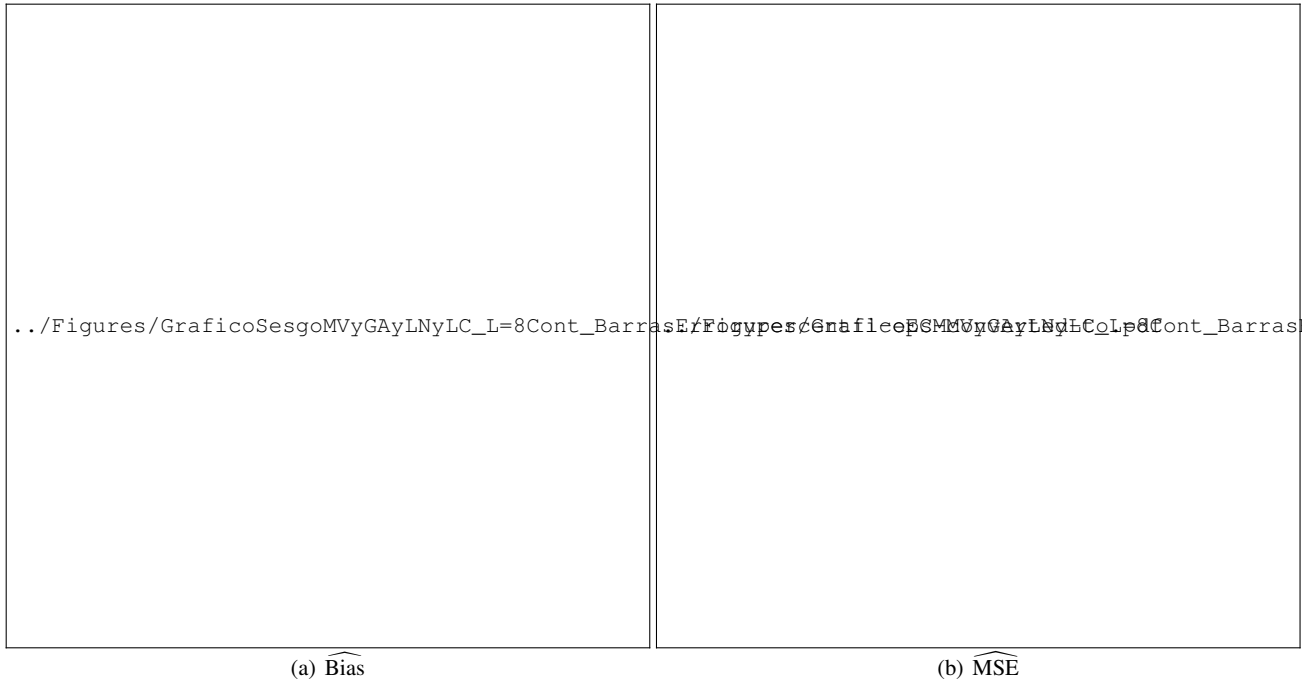
Color	$n$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_\Gamma$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
Yellow	16	-7.21 [-17.4; -2.47]	-4.34 [-11.5; -1.88]	-3.23 [-9.40; -1.64]	-6.74 [-14.5; -1.86]
	$n_b$	1392	1894	1926	1250
Red	16	-3.04 [-12.4; -1.72]	-3.42 [-10.3; -1.74]	-2.12 [-4.88; -1.43]	-3.27 [-12.0; -1.60]
	$n_b$	1924	1948	1988	1586
Blue	12	-4.62 [-15.5; -1.88]	-3.85 [-11.9; -1.68]	-2.35 [-7.46; -1.44]	-4.52 [-14.4; -1.59]
	$n_b$	1615	1879	1968	1376

**Table 7** Estimates of  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_\Gamma$ ,  $\hat{\alpha}_{LN}$  y  $\hat{\alpha}_{LC}$  in the yellow sample without the smallest observation, along with their bootstrap confidence intervals and effective bootstrap sample sizes ( $n_b$ ).

	$\hat{\alpha}_{ML}$	$\hat{\alpha}_\Gamma$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
	-5.73 [-16.9; -2.21]	-3.97 [-12.0; -1.80]	-3.13 [-10.2; -1.644]	-4.81 [-15.0; -1.70]
$n_b$	1538	1913	1940	1371

and so have  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_\Gamma$ . For  $n = 9$  none of the ML and LC methods converge, thus the value  $-20$ . As the sample size increases, the estimates stabilize, showing that the LN method is the most stable while  $\hat{\alpha}_{ML}$  and  $\hat{\alpha}_{LC}$  are the most unstable, giving poor estimates for small sample sizes. It can also be observed that  $n_b$  is smaller for MC and LC methods for small sample sizes, suggesting that these techniques are more prone to numerical instabilities in such cases.

In Table 9 we see the length of the confidence intervals depicted in Fig. 15. In most cases, the ML and LC produce wider intervals than those from other methods. All the widths decrease as the sample size increases; that is, estimates become progressively more precise. It is noteworthy that the shortest confidence interval corresponds to the LN method in most cases. It should be noted that there is no confidence interval for the cases where ML and LC methods did not converge.



**Fig. 12** Bias and MSE for contaminated data,  $\varepsilon = 0.05$  and  $L = 8$ .

**Table 8** Estimates of  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_T$ ,  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_{LC}$  for five nested samples from the ESAR image, along with their bootstrap confidence intervals and effective bootstrap sample sizes ( $n_b$ ).

Color	$n$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
Yellow	9	-20.00 - 857	-7.97 [-13.69; -1.83] 1642	-4.73 [-13.74; -1.63] 1740	-20.00 - 947
Red	25	-10.31 [-17.75; -3.02] 1332	-5.42 [-13.22; -2.28] 1900	-4.30 [-12.22; -2.02] 1957	-9.81 [-16.30; -2.16] 1159
Blue	49	-3.07 [-6.02; -2.08] 1997	-3.12 [-5.24; -1.98] 1999	-2.44 [-4.01; -1.73] 2000	-3.28 [-11.54; -1.89] 1893
Green	81	-2.24 [-3.20; -1.82] 2000	-2.15 [-2.79; -1.70] 2000	-2.03 [-2.65; -1.63] 2000	-2.45 [-4.46; -1.70] 1997
Magenta	121	-2.04 [-2.58; -1.73] 2000	-1.99 [-2.43; -1.66] 2000	-2.05 [-2.50; -1.67] 2000	-2.13 [-2.99; -1.67] 2000

It is important to clarify that this analysis is performed for small samples corresponding to the chosen windows sizes.

**Table 9** Length of bootstrap confidence intervals showed in Fig.15

$n$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
9	—	11.86	12.12	—
25	14.74	10.94	10.20	14.14
49	3.94	3.25	2.28	9.65
81	1.38	1.09	1.03	2.75
121	0.85	0.77	0.83	1.31

Figure 16 shows the same ESAR image with a corner reflector in a dark background. We selected nested samples to estimate  $\alpha$ ; only the two largest (in yellow and red) include the corner reflector. Table 10 shows the estimates, the confidence interval and the effective bootstrap sample size  $n_b$ . None of the estimators converge for  $n = 9$ . For the rest of the sample sizes, LN estimator is the most stable. Furthermore, the LN estimator does not show any change between the green sample and the yellow sample, which contain the corner reflector. All the other methods showed changes in the estimation by changing the type of texture from one sample



**Fig. 13** Samples used to estimate  $\hat{\alpha}$ .

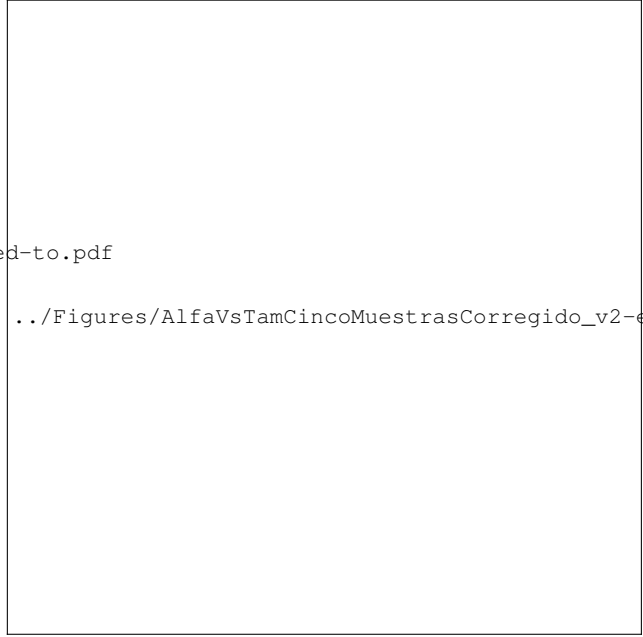


**Fig. 14** Five samples used to compare  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_\Gamma$ ,  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_{LC}$ .

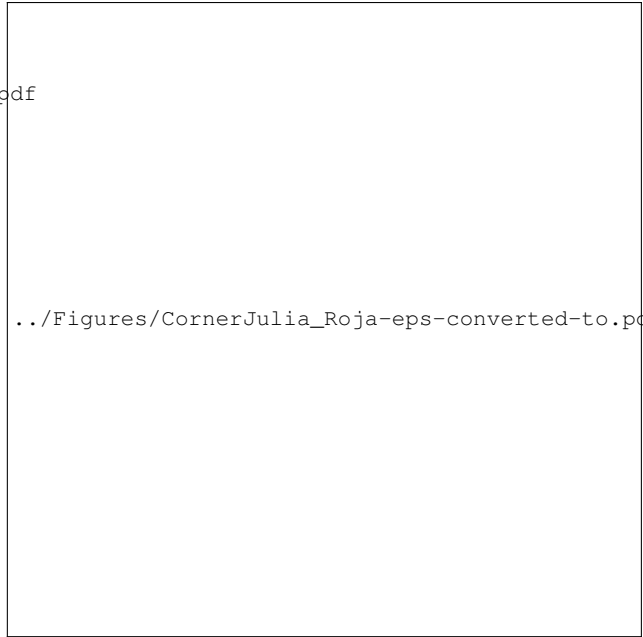
to another. That is the kind of stability sought for a good estimator for these kinds of problems of image processing and understanding. It is important to note that the confidence intervals are larger than those presented in Table 8, a likely effect of the corner reflector.

## 6 Conclusions

This work presented improvements to the MDE estimator for the texture parameter of the  $\mathcal{G}^0$  model for intensity and multilook data, with respect to the one studied in [28]. We assessed two asymmetric kernels: Lognormal and  $\Gamma$ , along with the LSCV method to obtain the bandwidth instead of using a fixed one empirically chosen. We also implemented an optimization procedure to find the minimum, maximum, and the roots of the functions associated with each of the estimators analyzed. We evaluated the kernels in the estimation of the underlying density function in terms of MISE and the percentage of non-convergence cases. We show that the IG



**Fig. 15**  $\hat{\alpha}_{ML}$ ,  $\hat{\alpha}_\Gamma$ ,  $\hat{\alpha}_{LN}$  and  $\hat{\alpha}_{LC}$  for corresponding samples to image 14.



**Fig. 16** Nested samples with and without a corner reflector.

kernel, with fixed bandwidth obtained by LSCV, has higher MISE in all of the studied cases and a higher percentage of non-convergence cases in most of the instances analyzed with respect to the one presented by the  $\Gamma$  and Lognormal kernels. For these reasons, we chose the latter to pursue the analysis.

A Monte Carlo study allowed us to verify that: a) On the one hand, the ML and LC methods have similar behavior; on the other, the MDE estimators using LN and  $\Gamma$  kernels for both uncontaminated and contaminated data perform well.

**Table 10** Estimated  $\hat{\alpha}$ , confidence intervals and effective bootstrap sample size ( $n_b$ ) in nested boxes.

Color	$n$	$\hat{\alpha}_{ML}$	$\hat{\alpha}_T$	$\hat{\alpha}_{LN}$	$\hat{\alpha}_{LC}$
Green	30	-13.04 [18.6; -3.56]	-6.82 [-14.7; -2.56]	-7.76 [-15.4; -2.5]	20.00 -
$n_b$		1131	1888	1816	1810
Yellow	81	-4.72 [-10.8; -2.97]	-5.09 [-9; -2.82]	-7.99 [-14.4; -3.41]	-4.29 [-13.7; -2.43]
$n_b$		1989	1992	1943	1893
Red	144	-6.94 [-15.3; -4.40]	-8.42 [-13.5; -4.12]	-8.53 [-13.9; -4.12]	-6.79 [-16.6; -3.28]
$n_b$		1956	1992	1943	1667

The SEIF function confirms this behavior. b) For uncontaminated and textured data, the MDE estimator is competitive with respect to the ML and LC in terms of bias and MSE. However, MDE outperforms ML and LC estimators in the presence of contamination, either using  $\Gamma$  or LN kernels, with the latter being the most stable. c) The percentage of non-convergence cases presented by MDE estimators is significantly larger than ML and LC estimators. d) The computational time of the MDE estimators is significantly larger than ML and LC estimators.

MDE, together with the Lognormal kernel, showed the best performance in the application to actual data, exceeding ML and LC estimators.

MDE estimators have excellent properties measured by its bias, mean square error, and the number of cases of non-convergence. They are competitive with the ML estimators and LC estimators in situations without contamination and outperform these classical methods in the presence of small contamination levels. MDE estimators have the best performance in the presence of a reflector corner.

For these reasons, we conclude that it is advisable to use  $\hat{\alpha}_T$  with the LN kernel, especially when using small samples and/or in the presence of contaminated data. Although the additional computational cost of this MDE estimator is not negligible, its advantages are more important than this increase in processing time.

All studies were made in the R platform and language for statistical computing [42].

## References

- Adelson, E., Anderson, C., Bergen, J., Burt, P., Ogden, J.: Pyramid methods in image processing. *RCA engineer* **29**(6), 33–41 (1984)
- Allende, H., Frery, A.C., Galbiati, J., Pizarro, L.: M-estimators with asymmetric influence functions: the  $\mathcal{G}_A^0$  distribution case. *Journal of Statistical Computation and Simulation* **76**(11), 941–956 (2006). DOI 10.1080/10629360600569154
- Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., Tuckey, J.: *Robust Estimates of Location: Survey and Advances*. New Jersey: Princeton University Press (1972)
- Anfinson, S.N., Eltoft, T.: Application of the matrix-variate Mellin transform to analysis of polarimetric radar images. *IEEE Transactions on Geoscience and Remote Sensing* **49**(6), 2281–2295 (2011). DOI 10.1109/TGRS.2010.2103945
- Aviyente, S., Ahmad, F., Amin, M.: Information theoretic measures for change detection in urban sensing applications. In: *IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, pp. 1–6 (2007)
- Beran, R.: Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**(3), 445–463 (1977). DOI 10.1214/aos/1176343842
- Bouezmarni, T., Scaillet, O.: Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* **21**(2), 390–412 (2005)
- Brisco, B., Schmitt, A., Murnaghan, K., Kaya, S., Roth, A.: SAR polarimetric change detection for flooded vegetation. *International Journal of Digital Earth* **6**(2), 103–114 (2013). DOI 10.1080/17538947.2011.608813
- Buckland, S.: Monte Carlo methods for confidence interval estimation using the bootstrap technique. *Journal of Applied Statistics* **10**(2), 194–212 (1983). DOI 10.1080/02664768300000017
- Bujor, F., Trouve, E., Valet, L., Nicolas, J.M., Rudant, J.P.: Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images. *IEEE Transactions on Geoscience and Remote Sensing* **42**(10), 2073–2084 (2004)
- Bustos, O.H., Lucini, M.M., Frery, A.C.: M-estimators of roughness and scale for  $\mathcal{G}_A^0$ -modelled SAR imagery. *EURASIP Journal on Applied Signal Processing* **1**(1), 105–114 (2002). DOI 10.1155/S1110865702000392
- Cao, R., Cuevas, A., Fraiman, R.: Minimum distance density-based estimation. *Computational Statistics & Data Analysis* **20**(6), 611–631 (1995)
- Carvalho, N., Sant’Anna Bins, L., Sant’Anna, S.J.: Analysis of stochastic distances and Wishart mixture models applied on Pol-SAR images. *Remote Sensing* **11**(24) (2019). DOI 10.3390/rs11242994
- Cassetti, J., Gambini, J., Frery, A.C.: Parameter estimation in SAR imagery using stochastic distances. In: *Proceedings of The 4th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR)*. Tsukuba, Japan (2013)
- Chan, D., Rey, A., Gambini, J., Frery, A.C.: Low-cost robust estimation for the single-look G10 model using the Pareto distribution. *IEEE Geoscience and Remote Sensing Letters* **17**(11), 1879–1883 (2020). DOI 10.1109/LGRS.2019.2956635
- Chen, S.: Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* **31**(2), 131–145 (1999)
- Chen, S.: Probability density functions estimation using Gamma kernels. *Annals of Institute Statistical Mathematics* **52**(3), 471–480 (2000)
- Cribari-Neto, F., Frery, A.C., Silva, M.F.: Improved estimation of clutter properties in speckled imagery. *Computational Statistics and Data Analysis* **40**(4), 801–824 (2002). DOI 10.1016/S0167-9473(02)00102-0

19. Csizsár, I.: Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* **2**, 299–318 (1967)
20. Davison, A., Hinkley, D.: *Bootstrap Methods and their Application*. Cambridge University Press (1997)
21. Dobélé-Kpoka, F.L.: Non parametric method of mixed associated kernels and applications. Ph.D. thesis, Université de Franche-Comté (2013). URL <https://tel.archives-ouvertes.fr/tel-01124288>
22. Efron, B., Tibshirani, R.: *An Introduction to the Bootstrap*. No. 57 in *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, Florida, USA (1993)
23. Fei, Y., Rong, G., Wang, B., Wang, W.: Parallel L-BFGS-B algorithm on GPU. *Computers & Graphics* **40**, 1 – 9 (2014). DOI <https://doi.org/10.1016/j.cag.2014.01.002>. URL <http://www.sciencedirect.com/science/article/pii/S0097849314000119>
24. Frery, A.C., Correia, A.H., Rennó, C.D., Freitas, C.C., Jacobo-Berlles, J., Mejail, M.E., Vasconcellos, K.L.P.: Models for synthetic aperture radar image analysis. *Resenhas (IME-USP)* **4**(1), 45–77 (1999)
25. Frery, A.C., Cribari-Neto, F., Souza, M.O.: Analysis of minute features in speckled imagery with maximum likelihood estimation. *EURASIP Journal on Advances in Signal Processing* **2004**(16), 2476–2491 (2004). DOI [10.1155/S111086570440907X](https://doi.org/10.1155/S111086570440907X)
26. Frery, A.C., Jacobo-Berlles, J., Gambini, J., Mejail, M.: Polarimetric SAR image segmentation with B-Splines and a new statistical model. *Multidimensional Systems and Signal Processing* **21**, 319–342 (2010). DOI [10.1007/s11045-010-0113-4](https://doi.org/10.1007/s11045-010-0113-4)
27. Frery, A.C., Müller, H.J., Yanasse, C.C., Sant’Anna, S.J.: A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing* **35**(3), 648–659 (1997). DOI [10.1109/36.581981](https://doi.org/10.1109/36.581981)
28. Gambini, J., Casseti, J., Lucini, M., Frery, A.: Parameter estimation in SAR imagery using stochastic distances and asymmetric kernel. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **8**, 365–375 (2015)
29. Gao, G.: Statistical modeling of SAR images: A survey. *Sensors* **10**(1), 775–795 (2010)
30. Horn, R.: The DLR airborne SAR project E-SAR. In: *Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1624–1628. IEEE Press (1996). DOI [10.1109/IGARSS.1996.516751](https://doi.org/10.1109/IGARSS.1996.516751)
31. Jin, X., Kawczak, J.: Birnbaum-Saunders and Lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance* **4**(1), 103–124 (2003). URL <http://EconPapers.repec.org/RePEc:cuf:journl:y:2003:v:4:i:1:p:103-124>
32. Khan, S., Guida, R.: Application of Mellin-Kind statistics to polarimetric  $\mathcal{G}$  distribution for SAR data. *IEEE Transactions on Geoscience and Remote Sensing* **52**(6), 3513–3528 (2014)
33. Luenberger, D., Ye, Y.: *Linear and Nonlinear Programming*. Springer Science+Business Media, LLC (2008)
34. Mejail, M., Jacobo-Berlles, J.C., Frery, A.C., Bustos, O.H.: Classification of SAR images using a general and tractable multiplicative model. *International Journal of Remote Sensing* **24**(18), 3565–3582 (2003)
35. Mejail, M.E., Frery, A.C., Jacobo-Berlles, J., Bustos, O.H.: Approximation of distributions for SAR images: Proposal, evaluation and practical consequences. *Latin American Applied Research* **31**, 83–92 (2001)
36. Nascimento, A.D., Cintra, R.J., Frery, A.C.: Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing* **48**(1), 373–385 (2010). DOI [10.1109/TGRS.2009.2025498](https://doi.org/10.1109/TGRS.2009.2025498)
37. Nicolas, J.M., Anfinson, S.N.: Introduction to second kind statistics: Application of log-moments and log-cumulants to the analysis of radar image distributions. *Traitement du Signal* **19**(3), 139–167 (2002)
38. Nogueira, F., Marques, R., Medeiros, F.: SAR image segmentation based on unsupervised classification of log-cumulants estimates. *IEEE Geoscience and Remote Sensing Letters* pp. 1–3 (2019). DOI [10.1109/LGRS.2019.2941075](https://doi.org/10.1109/LGRS.2019.2941075)
39. Parr, W.C., Schucany, W.R.: Minimum distance estimation and components of goodness-of-fit statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**(2), 178–189 (1982)
40. Parzen, E.: On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**(3), 1065–1076 (1962)
41. Penna, P., Mascarenhas, N.: SAR speckle nonlocal filtering with statistical modeling of Haar wavelet coefficients and stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing* **57**(9), 7194–7208 (2019). DOI [10.1109/TGRS.2019.2912153](https://doi.org/10.1109/TGRS.2019.2912153)
42. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2016). URL <https://www.R-project.org/>
43. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* **27**(3), 832–837 (1956)
44. Rousseeuw, P.J., Verboven, S.: Robust estimation in very small samples. *Computational Statistics and Data Analysis* **40**(4), 741–758 (2002)
45. Rudemo, M.: Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics* **9**(2), 65–78 (1982)
46. Salicrú, M., Morales, D., Menendez, M.L., Pardo, L.: On the applications of divergence type measures in testing statistical hypotheses. *Journal of Multivariate Analysis* **51**(2), 372–391 (1994)
47. Scailliet, O.: Density estimation using inverse and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics* **16**(1-2), 217–226 (2004)
48. Silva, M., Cribari-Neto, F., Frery, A.C.: Improved likelihood inference for the roughness parameter of the GA0 distribution. *Environmetrics* **19**(4), 347–368 (2008)
49. Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*, first edn. London: Chapman and Hall (1986)
50. Taneja, I.: On generalized information measures and their applications. In: P.W. Hawkes (ed.) *Advances in Electronics and Electron Physics*, vol. 76, pp. 327–413. Academic Press (1989). DOI [https://doi.org/10.1016/S0065-2539\(08\)60580-6](https://doi.org/10.1016/S0065-2539(08)60580-6)
51. Tison, C., Nicolas, J.M., Tupin, F., Maitre, H.: A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing* **42**(10), 2046–2057 (2004)
52. Vasconcellos, K.L.P., Frery, A.C., Silva, L.B.: Improving estimation in speckled imagery. *Computational Statistics* **20**(3), 503–519 (2005). DOI [10.1007/BF02741311](https://doi.org/10.1007/BF02741311)
53. White, R.G.: Change detection in SAR imagery. *International Journal of Remote Sensing* **12**(2), 339–360 (1991). DOI [10.1080/01431169108929656](https://doi.org/10.1080/01431169108929656)
54. Wolfowitz, J.: Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics* **5**, 9–23 (1953). DOI [10.1007/BF02949797](https://doi.org/10.1007/BF02949797)
55. Wolfowitz, J.: The minimum distance method. *The Annals of Mathematical Statistics* **28**(1), 75–88 (1957). DOI [10.1214/aoms/1177707038](https://doi.org/10.1214/aoms/1177707038)
56. Wu, C.: A cross-validation bandwidth choice for kernel density estimates with selection biased data. *Journal of Multivariate Analysis* **61**(1), 38 – 60 (1997). DOI <https://doi.org/10.1006/jmva.1997.1659>. URL <http://www.sciencedirect.com/science/article/pii/S0047259X97916596>
57. Yang, Y., Chen, Q., Xu, Q., Zhang, Y., Yong, Q., Liu, G.: Coseismic surface deformation of the 2014 Napa earthquake mapped by Sentinel-1A SAR and accuracy assessment with COSMO-SkyMed and GPS data as cross validation. *International Journal of Digital Earth* **10**(12), 1197–1213 (2017). DOI [10.1080/17538947.2017.1299806](https://doi.org/10.1080/17538947.2017.1299806)



58. Yue, D.X., Xu, F., Frery, A.C., Jin, Y.Q.: A generalized Gaussian coherent scatterer model for correlated SAR texture. *IEEE Transactions on Geoscience and Remote Sensing* **58**(4), 2947–2964 (2020). DOI 10.1109/TGRS.2019.2958125
59. Yue, D.X., Xu, F., Frery, A.C., Jin, Y.Q.: SAR image statistical modeling Part I: Single-pixel statistical models. *IEEE Geoscience and Remote Sensing Magazine* **9**(1), 82–114 (2021). DOI 10.1109/MGRS.2020.3004508